JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 52, No. 3, June 2017, pp. 1211–1245
COPYRIGHT 2017, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S002210901700031X

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

Firm Default Prediction: A Bayesian Model-Averaging Approach

Jeffrey Traczynski*

Abstract

I develop a new predictive approach using Bayesian model averaging to account for incomplete knowledge of the true model behind corporate default and bankruptcy filing. I find that uncertainty over the correct model is empirically large, with far fewer variables being significant predictors of default compared with conventional approaches. Only the ratio of total liabilities to total assets and the volatility of market returns are robust default predictors in the overall sample and individual industry groups. Model-averaged forecasts that aggregate information across models or allow for industry-specific effects substantially outperform individual models.

I. Introduction

Bankruptcy and default prediction is of interest to the creditors, customers, or suppliers of any firm as well as to policymakers and current and potential investors. Financial institutions require accurate assessments of a firm's future prospects, including the risk of bankruptcy or default, to price firm assets and credit derivatives. The latter has become particularly important after the prominent role of counterparty risk in the recent financial crisis. Studies have used many different variables as predictors to generate precise forecasts, find the variables that best serve as leading indicators of impending bankruptcy or default, and test the implications of theoretical models. The standard procedure in this literature is to declare a variable to be an important predictor if its parameter estimate is statistically significantly different from 0 and perform out-of-sample forecasting exercises.

The conventional approach has several shortcomings. Finding a set of variables that strongly predict bankruptcy or default could focus empirical research on refining the most important variables and discipline theoretical work.

*Traczynski (corresponding author), jtraczyn@hawaii.edu, College of Social Sciences, University of Hawaii at Manoa. I thank Hendrik Bessembinder (the editor), Steven Durlauf, Jane Cooley Fruehwirth, Bruce Hansen, S. Ghon Rhee, Karl Scholz, Tyler Shumway (the referee), Chris Taber, and participants in the University of Wisconsin–Madison Public Finance and Shidler College of Business seminars for their useful comments and suggestions. Special thanks go to Bill Taylor at the Center for High Throughput Computing at the University of Wisconsin–Madison for his invaluable assistance with computational tasks and to Ed Altman for providing data. All remaining errors are my own.

Unfortunately, there is no clear consensus in the literature on which variables are good predictors arising either from theory or empirics. For example, the canonical model of Merton (1974) proposes that firm default is a function of the value of firm assets and debts and the volatility of firm asset values, yet Bharath and Shumway (2008) and Campbell, Hilscher, and Szilagyi ((2008), p. 2901) find that the Merton distance-to-default measure "adds relatively little explanatory power to the reduced form variables already included" in the more atheoretical models of Shumway (2001) and Chava and Jarrow (2004). However, whether a variable adds explanatory power depends on the statistical model used and the assumptions underlying that model. Lack of knowledge of the true model can also lead to lower out-of-sample predictive accuracy, because ignoring this uncertainty leads to overconfidence in predictions from models that may not be correct. Collectively, these problems represent *model uncertainty* as analyzed in the asset pricing literature by Pastor and Stambaugh (1999), (2000), Pastor (2000), Avramov (2002), and Cremers (2002), among others.

Model uncertainty has received surprisingly little attention in bankruptcy and default prediction, despite evidence of its existence from the earliest to the most recent studies. Altman ((1968), p. 590) considers 22 potential covariates before settling on the 5 that comprise the *Z*-score, noting that "every [previous] study cited a different ratio as being the most effective." Similarly, Campbell et al. ((2008), p. 2902) point out that the current literature "varies in the choice of variables to predict bankruptcy and the methodology used." Tables 1 and 2 define a number of variables popular in the bankruptcy-prediction literature, beginning with Altman (1968) and Ohlson (1980). The differences in explanatory variables and the combinations in which they are used across papers show the disagreement over which covariates should be used to predict a firm's probability of filing for bankruptcy or defaulting and demonstrate that model uncertainty is prevalent in the current literature.

TABLE 1
Variable Description

Table 1 lists independent variables used in estimation and their descriptions. "Compustat" shows construction of variables from the Compustat database.

Variable	Description	Compustat
WC/TA	Working capital/Total assets	WCAPQ/ATQ
RE/TA	Retained earnings/Total assets	REQ/ATQ
EBIT/TA	Earnings before interest and taxes/Total assets	OIADPQ/ATQ
ME/TL	Market equity/Total liabilities	ME/LTQ
S/TA	Sales/Total assets	SALEQ/ATQ
TL/TA	Total liabilities/Total assets	LTQ/ATQ
CA/CL	Current assets/Current liabilities	ACTQ/LCTQ
NI/TA	Net income/Total assets	NIQ/ATQ
π_{MERTON}	Distance to default	—
SIGMA	Standard deviation of firm stock returns	—
AGE	log (Firm age in months)	DATADATE − BEGDAT
RSIZE	ln(Market equity/Total S&P 500 market value)	—
EXRET	ln(1 + Firm stock return) − log(1 + S&P 500 return)	—
CASH/MTA	Cash and short-term investments/(Market equity + Total liabilities)	CHEQ/(ME + LTQ)
MB	Market equity/Book equity	—
PRICE	ln(Firm stock price per share)	—
$1/\sigma_E$	1/Annualized volatility of firm stock returns	—
ME	Market equity	ME
F	Face value of firm debt	DLCQ+0.5 × DLTTQ

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 2
Variable Usage in Selected Papers

In Table 2, an "X" in a column indicates that the variable listed at the top is used as an explanatory variable for predicting firm bankruptcy in the paper at left. Variable definitions are provided in Table 1. The version of CASH/MTA used by Giordani, Jacobson, von Schedvin, and Villani (2014) excludes market equity, because it is not available in their data.

Paper	WC/TA	RE/TA	EBIT/TA	ME/TL	S/TA	TL/TA	CA/CL	NI/TA	π_{MERTON}	SIGMA	AGE	RSIZE	EXRET	CASH/MTA	MB	PRICE	$1/\sigma_E$	ME	F
Altman (1968)	X	X	X	X	X														
Ohlson (1980)	X					X	X	X											
Zimjewski (1984)						X	X	X											
Dichev (1998)	X	X	X	X	X	X	X	X							X			X	
Shumway (2001)	X	X	X	X	X	X	X	X		X	X	X	X						
Crosbie and Bohn (2003)									X										
Chava and Jarrow (2004)	X	X	X	X	X	X	X	X		X		X	X						
Hillegeist, Keating, Cram, and Lundstedt (2004)	X	X	X	X	X	X	X	X	X										
Vassalou and Xing (2004)									X				X		X			X	
Duffie, Saita, and Wang (2007)									X			X	X						
Bharath and Shumway (2008)								X	X				X				X	X	X
Campbell et al. (2008)						X		X	X	X		X	X	X	X	X			
Duffie, Eckner, Horel, and Saita (2009)									X			X	X						
Topaloglu and Yildirim (2009)	X	X	X	X	X	X	X	X	X	X		X	X						
Harada, Ito, and Takahashi (2010)									X										
Giordani et al. (2014)			X			X					X			X					

This paper makes several contributions to the study of firm bankruptcy and default. First, I address model uncertainty problems by developing a Bayesian model-averaging approach to analyzing firm bankruptcy and default predictability. Techniques incorporating uncertainty in linear models do not immediately translate to nonlinear hazard models.[1] I provide methods to extend model and parameter uncertainty analysis to such problems as bankruptcy or default prediction with limited dependent variables. I also allow for exchangeability uncertainty, where all observations are not generated by the same statistical model. Specifically, I allow predictive models to differ across industry groups, as in Chava and Jarrow (2004).

Researchers often avoid Bayesian approaches because of their computational cost. Bayesian model averaging of hazard models is particularly challenging because, unlike with linear models, there is no closed-form expression for a model's posterior likelihood when using standard parameter priors. I address both of these problems by using fully exponential Laplace approximations to high dimensional integrals as an accurate and computationally feasible solution. The Laplace approximations allow the Bayesian model-averaging approach to be applied easily to any setting with a limited dependent variable, not just hazard models.

Second, I apply this approach to investigate which covariates are and are not robust correlates of bankruptcy or default in different industry groups and across all firms. After accounting for model uncertainty, I find that only the ratio of total liabilities to total assets and the inverse of the annualized volatility of firm market equity are robust predictors of bankruptcy and default and that models using only these two variables better predict bankruptcies and defaults for all firms than models using all available covariates. These variables are similar to core

[1] See Shumway (2001) and Chava and Jarrow (2004) for discrete time-hazard models and Duffie, Saita, and Wang (2007) for a continuous time-hazard model.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

elements of the Merton (1974) default model, providing empirical support for the parsimony of Merton's theoretical model. Interestingly, the estimated probability of default from Merton's model is not a robust correlate, a finding similar to that of Bharath and Shumway (2008). I also identify variables that the data suggest not to use for prediction. These results can guide future researchers in variable selection for predictive models.

Third, I show that in out-of-sample forecasting, the model-averaged forecast is more accurate than a variety of other plausible models, including the model used by Campbell et al. (2008).[2] The magnitude of improvement is comparable to that from the inclusion of market variables as bankruptcy predictors noted by Shumway (2001). Model averaging resolves the problem in the literature, which is that including industry effects does not improve out-of-sample prediction accuracy, as shown by Chava and Jarrow (2004).

Fourth, this paper contributes to the model uncertainty literature more broadly by utilizing a generalized form of diluted model priors. A dilution prior offers a way for researchers to account for highly correlated covariates among the potential predictors by minimizing the effect of parameters estimated in the presence of multicollinearity on the final results. The dilution prior also produces slightly more accurate out-of-sample forecasts than the uniform prior used by Kandel and Stambaugh (1996), among others.

I find that model uncertainty is quantitatively large; most variables that appear to be significant predictors under conventional approaches are not strongly correlated with firm bankruptcy or default after accounting for model uncertainty. Although 15 variables meet standard statistical significance levels when using a hazard model to predict default over all firms, only 2 are robust to model uncertainty. The set of robust predictors of bankruptcy and default varies only slightly across industry groups, suggesting that exchangeability uncertainty affects parameter significance and prediction accuracy less than model uncertainty and that pooling all firms in a single sample is not a large source of parameter estimate instability. In out-of-sample forecasting, I find that the model-averaged forecast of defaults creates accuracy gains of nearly 4% per firm in the overall sample and between 1.2% and 5% per firm across industry groups compared with standard hazard models, a magnitude comparable to the inclusion of market variables in Shumway (2001). Campbell et al. (2008) show that firms differing by 1% in the distribution of predicted probabilities of firm failure, particularly at the extremes of the distribution, have large differences in firm stock returns. This finding suggests that improvements of 1.2% to 5% per firm in prediction accuracy are important for assessing the impact of default risk on other outcomes of interest. Thus, the model-averaging approach shows that model uncertainty in default prediction is empirically large and greatly affects which variables appear to be statistically significant and that model-averaged forecasts produce economically significant gains in out-of-sample predictive performance.

[2]Papers using the Bayesian model averaging of linear regression models to construct forecasts include those by Fernandez, Ley, and Steel (2001b) on cross-country growth, Avramov (2002) and Cremers (2002) on stock returns, Koop and Potter (2003) on U.S. gross domestic product growth, Stock and Watson (2005) on various macroeconomic time series, and Wright (2008), (2009) on exchange rates and U.S. inflation.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

II. Data

A. Description and Variable Creation

The variables defined in Table 1 are the entire set under consideration in this paper's empirical work. The model space is all models that can be made from combinations of these variables. With 19 explanatory variables, there are $2^{19} = 524{,}288$ models. Table 2 shows previous papers that have used these covariates as predictors of firm bankruptcy or default.[3]

I limit the sample to firms first publicly traded on or after Jan. 1, 1987. I obtain accounting data from Compustat Fundamentals Quarterly files and monthly and daily stock price data from the Center for Research in Security Prices (CRSP) from Jan. 1987 to Dec. 2008. I lag accounting variables from Compustat by 1 quarter so that all data are observable to the market at the start of the month. The final data set consists of monthly firm observations, as in Chava and Jarrow (2004). Table 1 describes all explanatory variables and the Compustat data series I use to construct them. I obtain data on bankruptcy filings of publicly traded companies from daily reports of U.S. bankruptcy courts from Jan. 1987 to Dec. 2009, as compiled by New Generation Research. I consider a firm to be bankrupt as of the date of filing for either Chapter 7 or Chapter 11 bankruptcy. If a firm files for bankruptcy more than once, the first filing is the date of bankruptcy.

I obtain data on firm defaults over the same period from the Altman–Kuehne/New York University (NYU) Salomon Center Database. I define a default as either an event recorded in the Altman database or a bankruptcy filing, although this may undercount the true number of defaults if some of the firms covered by the Compustat database are not covered in the Altman database. If a firm has more than one default, I consider the first default to be the date of default.

The first five variables in Table 1 are the components of the Z-score created by Altman (1968). These are working capital to total assets (WC/TA), retained earnings to total assets (RE/TA), earnings before interest and taxes to total assets (EBIT/TA), sales to total assets (S/TA), and market equity to total liabilities (ME/TL). The three accounting variables from Ohlson (1980) and Zimjewski (1984) are the ratios of net income to total assets (NI/TA), total liabilities to total assets (TL/TA), and current assets to current liabilities (CA/CL). π_{MERTON} is an estimated probability of default based on the structural model of firm default in Merton (1974). I calculate the distance to default following the iterative procedure described by Vassalou and Xing (2004).

The next seven variables are market-based. SIGMA is the idiosyncratic standard deviation of a firm's stock returns and measures the variability of a firm's cash flows. I calculate SIGMA for each month by regressing the monthly returns of a firm's stock over the previous 12 months on the monthly value-weighted

[3]Some papers use additional variables not considered here. These are either slight modifications of other variables included in Table 1 or yearly macroeconomic variables that are controlled for through the flexible baseline hazard described in Section III.B. When there are two similar variables, I include the one appearing first in the literature. For example, Campbell et al. (2008) create different versions of net income to total assets (NI/TA) and total liabilities to total assets (TL/TA). Their measures have correlations between 0.8 and 0.94 with traditional NI/TA and TL/TA in this sample. I use the Campbell et al. (2008) versions in the CHS models in forecasting exercises later in the paper.

Standard & Poor's (S&P) 500 index return over the same period. SIGMA is the standard deviation of the summed residuals of this regression.[4] AGE is the natural log of the number of months since the firm first became publicly traded, as recorded in the CRSP data. RSIZE measures the relative size of the market value of the firm's equity to the market value of the entire S&P 500, and EXRET measures the excess return on the firm's stock relative to the returns on the value-weighted S&P 500 index. CASH/MTA is the ratio of the firm's short-term assets to the market value of all assets, capturing the firm's liquidity. MB is the ratio of market equity to book equity, and PRICE is the natural log of the firm's stock price. I construct firm book equity as in Cohen, Polk, and Vuolteenaho (2003).

The final three variables are either proxies for inputs or actual inputs into π_{MERTON}. $1/\sigma_E$, the inverse of the annualized volatility of market equity, is a proxy for the volatility of firm assets $1/\sigma_A$, whereas market equity (ME) and the face value of debt, F, directly enter the calculation of π_{MERTON}. These variables are used by Bharath and Shumway (2008) to evaluate the predictive power of π_{MERTON} when its component variables are also included in the model.

For a firm-month to appear in the data, all 19 explanatory variables must be observed. All predictions are at a 12-month horizon, as is common in the literature.[5] Many variables feature a small number of extreme values. To limit the influence of outliers and follow conventions, I winsorize all variables at the 1st and 99th percentiles of their pooled observations except π_{MERTON}, AGE, and PRICE. π_{MERTON} is bounded between 0 and 1. Because the sample is limited to firms that first became publicly traded on or after Jan. 1, 1987, AGE is winsorized at this level. PRICE is winsorized above \$15 per share, as in Campbell et al. (2008).

B. Summary Statistics

I present summary statistics in Table 3 for both the full sample of firms in Panel A and a subsample of firms in the month in which they declared bankruptcy in Panel B. All values in Table 3 are reported after winsorization. The statistics presented in Panels A and B reflect the intuition that bankrupt firms have higher debt, lower asset values, and more volatile income streams, and the differences in means and medians across the two panels suggest that all of these variables may help predict bankruptcy.

To investigate whether different covariates might have differing predictive power for forecasting bankruptcies or defaults in different industries, I divide the firms into subsamples based on Standard Industrial Classification (SIC) codes available from CRSP and Compustat.[6] I then present results for the four largest industry groups: manufacturing (SIC codes 2000–3999); transportation, communications, and utilities (codes 4000–4999); retail trade (codes 5200–5999); and service industries (codes 7000–8899). This classification scheme is similar to that of Chava and Jarrow (2004).

[4] Value weighting is calculated by CRSP. SIGMA is considered missing if there are fewer than 6 monthly firm stock returns in the CRSP data over the preceding 12-month period.

[5] A firm is therefore considered censored in the data 12 months before filing. For example, for a firm that declares bankruptcy in Mar. 2005, I use data on and prior to Mar. 2004 to form predictions.

[6] Every firm-month is classified by its SIC code in that month, so a firm whose SIC code changes is classified in its new industry group as of the month of the SIC code change.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 3
Summary Statistics

Table 3 contains bankruptcy data from New Generation Research and firm data from CRSP/Compustat. *N* equals the total number of firm-months in each sample. Panel A includes all firm-months in the data regardless of bankruptcy status, whereas Panel B includes only firm-months in which the firm files for bankruptcy. ME and F are measured in millions of dollars. All variables are as defined in Section II.A and Table 1.

	Panel A. All Firms (N = 544,422)					*Panel B. Bankrupt Firms (N = 163)*				
Variable	Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
WC/TA	0.306	0.272	0.273	−0.389	0.879	−0.054	−0.045	0.278	−0.389	0.879
RE/TA	−0.514	−0.011	1.493	−8.294	0.823	−1.214	−0.344	2.152	−8.294	0.505
EBIT/TA	−0.008	0.011	0.075	−0.338	0.113	−0.057	−0.028	0.083	−0.338	0.059
ME/TL	9.833	2.794	18.525	0.033	107.742	0.691	0.075	3.235	0.033	36.323
S/TA	0.281	0.236	0.222	0.000	1.134	0.292	0.249	0.219	0.000	1.134
TL/TA	0.441	0.432	0.255	0.038	1.247	0.828	0.850	0.275	0.038	1.247
CA/CL	3.519	2.153	3.983	0.271	24.089	1.272	0.895	2.138	0.271	24.089
NI/TA	−0.021	0.004	0.085	−0.440	0.094	−0.103	−0.059	0.121	−0.440	0.047
π_{MERTON}	0.071	0.000	0.212	0.000	1.000	0.707	0.955	0.376	0.000	1.000
SIGMA	0.162	0.139	0.099	0.029	0.551	0.268	0.236	0.120	0.057	0.551
AGE	3.861	3.989	0.882	1.792	5.557	4.021	4.043	0.709	2.197	5.231
RSIZE	−10.972	−11.095	1.775	−14.933	−5.521	−13.132	−13.277	1.406	−14.933	−8.896
EXRET	−0.015	−0.014	0.174	−0.543	0.488	−0.258	−0.268	0.279	−0.543	0.488
CASH/MTA	0.125	0.064	0.167	0.000	0.889	0.073	0.028	0.135	0.000	0.889
MB	3.132	2.138	7.145	0.063	25.112	6.613	0.827	9.191	0.063	25.112
PRICE	1.857	2.106	0.975	−4.159	2.708	−0.343	−0.402	1.177	−3.219	2.708
$1/\sigma_E$	1.726	1.482	0.924	0.453	6.523	0.867	0.748	0.407	0.453	2.685
ME	7.235	1.075	23.628	0.013	249.481	0.499	0.120	1.523	0.013	15.495
F	1.412	0.051	5.472	0.000	51.670	3.054	0.704	6.542	0.000	51.670

Table 4 reports summary statistics by industry group. Manufacturing and service industry firms appear similar in observables, whereas firms in the transportation and retail industry groups have higher market equity and debt and are more heavily leveraged. As a result, many of the accounting ratios are smaller in absolute value for transportation and retail firms. The difference in leverage is reflected in considerably higher π_{MERTON} values for transportation and retail firms than for manufacturing and service companies. The means of SIGMA and $1/\sigma_E$ show that the transportation and retail firms have lower volatility of market equity.

Table 5 lists the number of firms in each year that file for bankruptcy or default in the following year, as I predict bankruptcies and defaults at a 12-month horizon. There are fewer bankruptcies in these data than in the data used by Campbell et al. (2008), because I require more variables to be observable for a firm to remain in the data. The percentage of firms in the data declaring bankruptcy or defaulting by year is generally similar for the overlapping years.

Table 6 shows the number of firms, bankruptcies, and defaults by industry group. Manufacturing firms are the largest industry group in this sample, but retail firms have the highest bankruptcy rate, with nearly 18% of firms filing. Although differences in bankruptcy or default rates across industries do not necessarily imply that the determinants of bankruptcy or default differ across industries, there is substantial variation in bankruptcy and default rates across industry groups. Only 7.43% of service firms and 8.19% of manufacturing firms declare bankruptcy, with transportation firms ranking in the middle at 14.38%. The default rate is highest among transportation firms at 19.82%, with retail firms closely behind at 19.09%. The manufacturing and service industry groups have default rates of approximately 9% and 8%, respectively. This variation may result from industry groups facing different shocks over this period or from bankruptcy or default

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 4
Summary Statistics by Industry Group

In Table 4, firm data are from CRSP/Compustat. *N* equals the total number of firm-months in each sample. ME and F are measured in millions of dollars. All variables are as defined in Section II.A and Table 1.

	Industry											
	Manufacturing			Transportation			Retail			Service		
Variable	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
WC/TA	0.394	0.384	0.263	0.097	0.048	0.206	0.224	0.222	0.228	0.289	0.268	0.273
RE/TA	−0.646	−0.008	1.659	−0.244	0.006	0.984	−0.049	0.098	0.776	−0.647	−0.061	1.620
EBIT/TA	−0.016	0.011	0.084	0.003	0.013	0.053	0.017	0.021	0.048	−0.009	0.011	0.075
ME/TL	12.422	3.970	21.280	3.617	1.040	10.064	5.435	1.953	10.715	10.144	4.129	17.273
S/TA	0.251	0.234	0.180	0.215	0.143	0.206	0.506	0.472	0.238	0.269	0.229	0.200
TL/TA	0.404	0.368	0.256	0.588	0.595	0.253	0.493	0.479	0.229	0.419	0.382	0.245
CA/CL	4.411	2.822	4.537	2.123	1.337	2.747	2.253	1.709	2.222	3.106	2.083	3.340
NI/TA	−0.028	0.005	0.092	−0.011	0.004	0.065	0.002	0.010	0.053	−0.024	0.005	0.090
π_{MERTON}	0.059	0.000	0.194	0.156	0.000	0.321	0.091	0.000	0.244	0.049	0.000	0.163
SIGMA	0.162	0.139	0.096	0.141	0.114	0.097	0.141	0.121	0.081	0.180	0.152	0.108
AGE	3.918	4.043	0.883	3.818	3.932	0.879	3.915	4.043	0.887	3.768	3.871	0.869
RSIZE	−11.011	−11.060	1.736	−10.596	−10.622	1.893	−10.832	−10.823	1.796	−11.012	−11.052	1.730
EXRET	−0.015	−0.013	0.173	−0.013	−0.010	0.160	−0.012	−0.010	0.155	−0.019	−0.014	0.189
CASH/MTA	0.135	0.073	0.172	0.099	0.044	0.154	0.067	0.033	0.097	0.152	0.090	0.180
MB	2.854	2.282	3.135	6.741	1.636	9.237	3.872	1.906	5.135	2.923	2.628	4.137
PRICE	1.856	2.140	0.946	2.047	2.536	0.926	2.017	2.398	0.874	1.799	2.079	1.005
$1/\sigma_E$	1.715	1.494	0.903	1.983	1.718	1.133	1.890	1.738	0.892	1.583	1.377	0.848
ME	6.599	1.083	22.018	12.312	1.782	36.434	7.603	1.295	23.332	6.621	1.176	20.943
F	1.117	0.036	4.531	5.399	0.572	11.959	1.047	0.118	3.519	0.630	0.019	2.730
N		257,576			51,189			44,435			133,338	

https://doi.org/10.1017/S0022109017000317X Published online by Cambridge University Press

TABLE 5
Bankruptcy Filings and Defaults by Year

Table 5 contains bankruptcy data from New Generation Research, default data from Altman–Kuehne/NYU Salomon Center Database, and firm data from CRSP/Compustat. "Bankruptcies" and "Defaults" count the number of firms in a given year that file for bankruptcy or default in the following year.

Year	Bankruptcies	Defaults	No. of Firms	Bankruptcy (%)	Default (%)
1987	2	2	517	0.39	0.39
1988	6	7	840	0.71	0.83
1989	4	4	1,044	0.38	0.38
1990	8	9	1,259	0.64	0.71
1991	11	11	1,411	0.78	0.78
1992	11	12	1,798	0.61	0.67
1993	8	9	2,198	0.36	0.41
1994	20	21	2,708	0.74	0.78
1995	23	26	2,988	0.77	0.87
1996	24	26	3,499	0.69	0.74
1997	40	46	3,840	1.04	1.20
1998	55	71	3,894	1.41	1.82
1999	70	85	3,667	1.91	2.32
2000	118	133	3,706	3.18	3.59
2001	90	104	3,558	2.53	2.92
2002	63	72	3,284	1.92	2.19
2003	24	28	3,035	0.79	0.92
2004	19	24	2,967	0.64	0.81
2005	17	19	2,923	0.58	0.65
2006	15	17	2,924	0.51	0.58
2007	36	41	2,854	1.26	1.44
2008	57	67	2,409	2.37	2.78

TABLE 6
Bankruptcies and Defaults by Industry Group

Table 6 contains bankruptcy data from New Generation Research, default data from Altman–Kuehne/NYU Salomon Center Database, and firm data and SIC codes from CRSP/Compustat. The numbers of bankruptcies, defaults, and firms are calculated over all available years.

		No. of				
SIC Code	Industry Name	Bankruptcies	Defaults	Firms	Bankruptcies/ Firms (%)	Defaults/ Firms (%)
2000–3999	Manufacturing	275	305	3,357	8.19	9.09
4000–4999	Transportation, communications, and utilities	111	153	772	14.38	19.82
5200–5999	Retail trade	95	101	529	17.96	19.09
7000–8899	Service industries	156	174	2,100	7.43	8.29

having different consequences for large and small firms, as the two industry groups with higher average market equity per firm show a higher percentage of firms filing for bankruptcy or defaulting.

III. Bayesian Model Averaging and Hazard Models[7]

A. Model-Averaged Parameter Estimates

Let $\hat{\beta}_m$ denote a parameter estimate obtained using model m, and let M denote the space of all possible models. Bayesian model averaging yields an estimate $\hat{\beta}_M$ calculated as

$$\hat{\beta}_M = \sum_{m \in M} \hat{\beta}_m \times P(m|y), \tag{1}$$

[7] In Section III, the term *model* refers to a particular combination of covariates used to estimate the hazard function.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

where $P(m|y)$ is the posterior probability that model m is true given data y. $P(m|y)$ is given by Bayes' rule

$$P(m|y) = \frac{P(m) \times P(y|m)}{\sum_{m \in M} P(m) \times P(y|m)}, \tag{2}$$

where $P(m)$ is the prior probability assigned to model m, and $P(y|m)$ is the marginal likelihood of the data given model m. $P(y|m)$ is given by the integral

$$P(y|m) = \int f(y|\beta_m, m) \times f(\beta_m|m) d\beta_m, \tag{3}$$

where $\beta_m = (\beta^1, \beta^2, \ldots)$ is a parameter vector, $f(y|\beta_m, m)$ is the likelihood of the data given the model m and the parameters β_m, and $f(\beta_m|m)$ is the prior distribution of β_m. To implement Bayesian model averaging, I now define the model and parameter priors and the likelihood function for the data.

B. Bayesian Estimation of Hazard Models

I compute estimates using a discrete-period hazard model with a nonparametric baseline hazard where each year has its own hazard rate for firm failure. This specification controls for year-specific shocks affecting all firms in the sample. The unit of observation is a firm-month, with model parameters estimated using a multiperiod logit over the pooled firm-month observations. The parameter estimates and variance–covariance matrix of a multiperiod logit estimated in this way are identical to those of a discrete-period hazard model.

Hazard model coefficients computed in this way are identified only up to scale within a model. To allow for comparison and averaging of coefficients across hazard models, I constrain the variance of the latent variable in the logit function to be 1 in every model. Adding this constraint fixes the scale of the coefficients and means that the coefficient on a given variable should be interpreted as the change in standard deviations of the latent variable associated with a 1-unit change in that variable. The likelihood function for the data is therefore a standard logistic likelihood function given by

$$\begin{aligned}
\ln f(y|\beta_m, m) &= \sum_i \sum_t \left(y_{it} \times \ln \left(\frac{1}{1 + e^{-\beta_m x_{mit}}} \right) \right. \\
&\quad \left. + (1 - y_{it}) \times \ln \left(\frac{e^{-\beta_m x_{it}}}{1 + e^{-\beta_m x_{mit}}} \right) \right), \\
y^*_{it} &= \beta_m x_{mit} + \epsilon_{it}, \\
y_{it} &= 1\left[y^*_{it} > 0\right], \\
\mathrm{var}\left(y^*_{it}\right) &= 1,
\end{aligned}$$

where y_{it} is an indicator equal to 1 if firm i declares bankruptcy in month t, y^*_{it} is the latent variable representing the firm's financial health, $\beta_m = (\beta^1, \beta^2, \ldots)$ is a parameter vector, x_{mit} is a set of explanatory variables in model m for firm i observable in month t, ϵ_{it} is an error term with a standard logistic distribution, and $1[y^*_{it} > 0]$ is an indicator function.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

For parameter priors, I use a separate prior formulation for the coefficients on the baseline hazard rates and for the coefficients on the covariates under analysis. Let $\beta_m = (\beta_b, \beta_c)$ denote the parameter vector for model m, where β_b represents the vector of coefficients on the baseline hazard rates, and β_c represents the coefficients on the covariates under analysis. A prior for the baseline hazard rates represents a prior belief about the average number of firm bankruptcies or defaults that might occur in each year. To make the prior as uninformative as possible, I use an improper flat prior for these parameters. Using an uninformative prior with year-specific baseline hazard rates is consistent with the frailty correlations in default described by Duffie et al. (2009). Thus, I assign as a prior distribution for β_b the improper prior

$$f(\beta_b) \quad \propto \quad 1.$$

The same approach to priors over model intercepts has been used in the context of ordinary least squares (OLS) models by Fernandez, Ley, and Steel (2001a), Fernandez et al. (2001b), and Ley and Steel (2009).

I assign as a prior distribution for each β_c the g-prior proposed by Zellner (1986) and used by Cremers (2002) and others, given by

$$f(\beta_c|m) \quad = \quad \mathrm{N}\left(0, g\left(X_m' \times X_m\right)^{-1}\right),$$

where N is a multivariate normal distribution of the same dimension as β_c, X_m is the centered matrix of covariates used in model m, and g is a scalar parameter. The prior for β_c is proper and centered at 0 in every dimension. Centering the prior at 0 for all variables shrinks posterior model parameter estimates toward 0, so the prior belief is that all variables are not useful predictors of bankruptcy or default.[8] The parameter g controls the relative weights put on the prior and data when forming the posterior distribution for each parameter vector β_m. I use the unit information prior recommended by Kass and Raftery (1995) and Fernandez et al. (2001a) by setting $g = 1/n$, where n is the sample size. This may be interpreted as the prior having as much effect on the posterior as one additional data point.[9]

The prior overall parameters of model m are the product of these two priors, given by

$$f(\beta_m|m) \quad = \quad f(\beta_c|m) \times f(\beta_b) \quad \propto \quad \mathrm{N}\left(0, g\left(X_m' \times X_m\right)^{-1}\right),$$

[8] See Stock and Watson (2005) for an interpretation of model-averaged estimates as shrinkage estimators.

[9] In this context, the importance of the choice of parameter prior mean is diminished by the sample size. In the smallest subsample analyzed, the prior mean accounts for approximately $1/44{,}500 \approx 0.00225\%$ of the model-averaged parameter. Even if a researcher had a strong prior belief that a variable should have a large effect, the prior mean would have minimal impact on the final estimate unless it was many orders of magnitude different from the observed effect of that variable in the data. Using the in-sample parameter maximum likelihood value as the prior mean has no effect on results. This may not be true in applications with fewer data available. See Fernandez et al. (2001a) for recommended values of g when n is small.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

and estimates $\hat{\beta}_m$ are posterior modes obtained by maximizing the posterior log likelihood

$$\begin{aligned} \hat{\beta}_m &= \underset{\beta_m}{\text{argmax}} \ln f(\beta_m|y,m) \\ &= \underset{\beta_m}{\text{argmax}} \{\ln f(\beta_m|m) + \ln f(y|\beta_m,m)\}. \end{aligned} \tag{4}$$

There is no closed-form expression for the posterior likelihood. I therefore use an iterated reweighted least squares algorithm to evaluate equation (4) numerically. I also compute variance estimates for the β_m parameters using the observed information matrix. I find H_m, the Hessian of the posterior likelihood function evaluated at $\hat{\beta}_m$, using the iterated reweighted least squares algorithm and set $\widehat{\text{var}}(\beta_m) = \text{diag}(H_m^{-1})$.

The mode of the posterior distribution is the most likely value of β_m given the data. I use the mode as the central characteristic of the posterior distribution, because it can be found with common routines and is faster to compute than the mean, which requires simulation via Monte Carlo methods. In practice, the computational time saved by using the mode instead of the mean is large. Additionally, the asymptotic normality of the posterior implies that the mode of the posterior distribution should be close to the mean in large samples.[10] I confirm this result using Monte Carlo simulation on subsamples. To the extent that the mode does not reflect a central tendency of the posterior distribution, the predictive accuracy of the Bayesian estimates will suffer relative to other models.

C. Laplace Approximation

With the model likelihood and parameter priors as defined previously, it is possible to find the marginal likelihood in equation (3). Without a closed form expression for the posterior likelihood function, this high dimensional integral must be evaluated directly. I therefore use the fully exponential Laplace approximation to the integral, so

$$P(y|m) = \frac{f(y|\hat{\beta}_m,m) \times f(\hat{\beta}_m|m)}{f(\hat{\beta}_m|y,m)},$$

where

$$f(\hat{\beta}_m|m) \approx |X_m' \times X_m|^{1/2}(2\pi g)^{-z/2} e^{\left(-1/2g\hat{\beta}_m' \times X_m' \times X_m \times \hat{\beta}_m\right)}$$

and

$$f(\hat{\beta}_m|y,m) \approx |H_m|^{1/2}(2\pi g)^{-(z+1)/2},$$

where z is the number of covariates included in model m. Tierney and Kadane (1986) and Tierney, Kass, and Kadane (1989) show that this approximation of $P(y|m)$ has an error of order $\mathcal{O}\left(n^{-2}\right)$, making it both accurate and easy to calculate. Clearly, the approximation becomes more accurate when there are more data available.

[10] For a simple proof of the asymptotic normality of the posterior, see Crain and Morgan (1975).

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

This approximation converges to equation (3) with probability 1 if the likelihood function in equation (4) is *Laplace regular*. Among other conditions, Laplace regularity requires that the integrals in equation (3) exist and are finite, the determinants of the Hessians are not 0 at their respective optima, and the log-likelihood functions have bounded partial derivatives for all parameters.[11] Crawford (1994) shows that any finite mixture of exponential family distributions has Laplace-regular log-likelihood functions when the distribution parameters are assumed to be identifiable. In the Appendix, I show that the integrand in equation (3) is an exponential family distribution. Common models of binary decisions, including logit and probit regressions, have likelihood functions in the exponential family of distributions, so this regularity assumption holds for many potential applications of these methods.

Other approximation methods are either less accurate, more computationally intensive, or both. The Bayesian information criterion (BIC) approximation to the posterior likelihood used by Volinsky, Madigan, Raftery, and Kronmal (1996) and the Akaike information criterion (AIC) approximation of Weakliem (1999) and others are approximations to the Laplace approximation, as shown for the BIC approximation by Raftery (1996). The runtime for the required maximum likelihood estimation is nearly identical to that for the Bayesian approach. Rosenkranz, Raftery, and Diehr (1994) and Azevedo-Filho and Shachter (1994) show that Monte Carlo approximation of the marginal likelihood requires around 20 times more computer time than the Laplace approximation, with a small 0.14% upper bound on accuracy gains.[12] The Laplace approximation is therefore a practical way to implement model averaging that removes the computing burden of Monte Carlo approximations, making Bayesian estimation much more computationally feasible for researchers across a wide range of problems.

D. Model Priors

Equation (2) shows the importance of model priors in the calculation of $P(m|y)$. The main results use a generalized form of the dilution priors suggested by George (1999) and Durlauf, Kourtellos, and Tan (2008), where the dilution prior $P_D(m)$ is given by

$$P_D(m) \quad \propto \quad |R_m| \times \prod_{j=1}^{J} p_j^{d_j} \left(1 - p_j\right)^{1-d_j},$$

where $|R_m|$ is the determinant of the correlation matrix of the explanatory variables in model m, J is the total number of candidate explanatory variables, p_j is the prior probability that $\beta^j \neq 0$, and d_j is an indicator for whether variable j is included in model m.[13]

[11] For formal discussion of the conditions of Laplace regularity, see Kass, Tierney, and Kadane (1990).

[12] As a robustness check, I compute the BIC and AIC approximations to $P(y|m)$ and include the details in the Appendix. Results are qualitatively similar. I also confirm the relative runtimes and accuracy gains from using Monte Carlo approximations in subsamples. The minimal accuracy gain from Monte Carlo estimation is a potential consequence of the sample size in this application.

[13] The form used by Durlauf et al. (2008) is based on the use of tree priors, where the relevant correlation matrix is the correlation matrix of variables in a given model that proxy for the same

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

This prior differs from the uniform prior of Kandel and Stambaugh (1996), Cremers (2002), and Avramov (2002) in the inclusion of the $|R_m|$ term. The dilution prior reflects the belief that covariates are imperfect empirical proxies for an underlying theoretical causal relationship, so models with multiple variables that proxy for the same causal mechanism should receive less weight. For example, if indebtedness causes bankruptcy filings, then a model with multiple measures of a firm's indebtedness receives a lower prior weight than a model with only one such covariate. The dilution prior can be thought of as an approximation to a prior that an experienced researcher might assign across models. The dilution prior penalizes model overfitting and minimizes the effect of parameters estimated in models with large multicollinearity on the final averaged parameter estimates. I set $p_j = 0.5$ to reflect a standard uniform prior, so all models receive equal prior weight except for the $|R_m|$ term. As a robustness check, I also calculate results using the uniform prior and a prior with $p_j = 5/19$ to reflect an expected model size of five covariates, as described by Sala-i-Martin, Doppelhofer, and Miller (2004).

Table 7 shows the cross-correlations between the variables described in Table 1 across the full sample of firms. There are three groups of variables with high within-group correlations: RE/TA, EBIT/TA, and NI/TA; WC/TA, ME/TL, TL/TA, CA/CL, and CASH/MTA; and RSIZE, PRICE, $1/\sigma_E$, and ME. The high correlations indicate that variables within a group are measuring the same fundamental characteristic of firms. The variables in the first group of RE/TA, EBIT/TA, and NI/TA are measures of firm income streams, whereas the variables in the second group of WC/TA, ME/TL, TL/TA, CA/CL, and CASH/MTA are measures of leverage and immediate access to operating money. The third group is a set of market variables reflecting changes in the firm's stock price. Models containing multiple variables from any one of these groups will have a lower prior weight because of the high correlations.

E. Model-Averaged Variance Estimates

Leamer ((1978), p. 118) shows that the variance of a model-averaged parameter β_M is

$$(5)\quad \widehat{\text{var}}(\beta_M|y) = \sum_{m\in M}\widehat{\text{var}}(\beta_m)\times P(m|y) + \sum_{m\in M}\left(\hat{\beta}_m - \hat{\beta}_M\right)^2 \times P(m|y),$$

where $\widehat{\text{var}}(\beta_m)$ is the estimated variance of parameter estimate $\hat{\beta}_m$ in model m. The first term in the model-averaged variance is directly analogous to equation (1), as it is the weighted sum of the estimated variances of the parameter estimates in different models, where the weights are the posterior probabilities of the corresponding models. As described previously, I estimate $\widehat{\text{var}}(\beta_m)$ as the diagonal elements of the inverse Hessian matrix evaluated at $\hat{\beta}_m$. The second term is the weighted sum of the squared deviations of each model's parameter estimate $\hat{\beta}_m$ from the model-averaged parameter estimate $\hat{\beta}_M$. Thus, the first term reflects within-model variance in estimates of $\hat{\beta}_m$, whereas the second term reflects between-model variance. The model-averaged standard errors are the square root of $\widehat{\text{var}}(\beta_M|y)$.

underlying causal theory rather than all variables in the model. See also Brock, Durlauf, and West (2003) for a discussion of the use of tree priors in Bayesian model averaging.

TABLE 7

Cross-Correlations

In Table 7, firm data are from CRSP/Compustat. All variables are as defined in Section II.A and Table 1. The number in each cell indicates the correlation between the row and column variables.

Variable	WC/TA	RE/TA	EBIT/TA	ME/TL	S/TA	TL/TA	CA/CL	NI/TA	π_{MERTON}	SIGMA	AGE	RSIZE	EXRET	CASH/MTA	MB	PRICE	$1/\sigma_E$	ME	F
WC/TA	1.000	−0.061	−0.116	0.451	−0.114	−0.623	0.684	−0.035	−0.256	0.081	−0.118	−0.027	−0.012	0.493	−0.125	0.068	−0.140	−0.088	−0.224
RE/TA		1.000	0.634	−0.176	0.208	−0.030	−0.064	0.598	0.023	−0.336	−0.046	0.257	0.020	−0.131	−0.132	0.401	0.289	0.107	0.088
EBIT/TA			1.000	−0.220	0.371	0.074	−0.172	0.878	0.014	−0.338	0.089	0.309	0.064	−0.209	−0.046	0.406	0.317	0.133	0.100
ME/TL				1.000	−0.231	−0.494	0.623	−0.158	−0.165	0.168	−0.142	0.116	0.057	0.043	−0.060	0.094	−0.115	0.038	−0.121
S/TA					1.000	0.241	−0.314	0.292	0.020	−0.090	0.031	−0.055	0.032	−0.271	−0.007	0.055	0.053	−0.047	−0.068
TL/TA						1.000	−0.536	−0.017	0.325	−0.053	0.103	0.019	0.015	−0.389	0.393	−0.050	0.099	0.039	0.230
CA/CL							1.000	−0.088	−0.159	0.088	−0.130	−0.023	−0.018	0.454	−0.065	0.009	−0.117	−0.061	−0.131
NI/TA								1.000	−0.023	−0.314	0.078	0.269	0.060	−0.145	−0.080	0.378	0.293	0.115	0.078
π_{MERTON}									1.000	0.106	0.025	−0.277	−0.037	−0.044	0.064	−0.224	−0.141	−0.086	0.172
SIGMA										1.000	−0.107	−0.324	0.001	0.105	0.020	−0.379	−0.610	−0.151	−0.160
AGE											1.000	0.057	0.044	−0.079	0.028	0.038	0.164	0.130	0.092
RSIZE												1.000	0.106	−0.208	0.067	0.722	0.485	0.540	0.316
EXRET													1.000	−0.096	0.011	0.152	0.048	0.039	0.015
CASH/MTA														1.000	−0.054	−0.202	−0.173	−0.099	−0.079
MB															1.000	0.012	0.003	0.009	0.078
PRICE																1.000	0.485	0.237	0.162
$1/\sigma_E$																	1.000	0.265	0.222
ME																		1.000	0.452
F																			1.000

https://doi.org/10.1017/S0022109017000317X Published online by Cambridge University Press

F. Variable Posterior Inclusion Probabilities

To determine which variables are most important for prediction, I calculate the posterior inclusion probability for each variable j as

$$P\left(\beta^j \neq 0|y\right) = \sum_{m \in M_j} P\left(m|y\right),$$

where $M_j = \{m|\beta^j \neq 0\}$. M_j is the set of all models that include variable j, and $P\left(\beta^j \neq 0|y\right)$ is the sum of the posterior probabilities of those models.

$P\left(\beta^j \neq 0|y\right)$ gives the probability that variable j is in the true model of firm bankruptcy or default. The interpretation of $P\left(\beta^j \neq 0|y\right)$ is different from that of a standard t-test for parameter significance. If a t-test on a coefficient estimate fails to reject the null hypothesis $H_0 : \beta^j = 0$, this cannot be properly interpreted as variable j's having no effect on the outcome of interest, only that the regression has not produced evidence that the effect is not 0. A t-test cannot offer conclusive evidence in favor of a null hypothesis.[14] However, if $P\left(\beta^j \neq 0|y\right)$ is close to 0, then this can be interpreted as the data indicating that variable j is not important. If $P\left(\beta^j \neq 0|y\right)$ is close to p_j, the prior probability that variable j is in the true model, then the data do not reveal much about the importance of variable j. A value of $P\left(\beta^j \neq 0|y\right)$ close to 1 means that the data provide strong evidence in favor of including variable j in the model of bankruptcy or default. The ability to interpret posterior inclusion probabilities in this way is a strength of Bayesian model averaging over traditional t-tests.

I use posterior inclusion probabilities to determine the set of variables most important in predicting firm bankruptcies and defaults. I also present the model-averaged coefficients and standard errors to compare between the conclusions drawn from posterior inclusion probabilities and those from hypothesis testing with t-tests that account for model uncertainty.

IV. Results

A. Bayesian Model-Averaging Results

Table 8 reports the model-averaged parameter estimates, standard errors, and posterior variable inclusion probabilities at a prediction horizon of 12 months for bankruptcies in the sample of transportation firms. Each set of estimates requires averaging results from $2^{19} = 524{,}288$ hazard models. In Table 8, estimate set 1 uses the dilution prior described in Section III.D, set 2 uses a uniform prior where all variables have a prior inclusion probability of 0.5, and set 3 uses a prior inclusion probability for each variable of 5/19 for an expected model size of five variables.[15] Following previous literature on Bayesian model averaging, a variable is a robust predictor of the outcome if its posterior inclusion probability is above 0.9, and

[14] Freedman (2009) shows that t-tests in hazard models have little power against general alternatives.

[15] In all industry groups, unreported results using the AIC or BIC in place of the posterior likelihood for model weighting are qualitatively similar for all three priors. See the Appendix for a discussion of the construction of these approximations to the Bayesian methods described previously.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 8

Bankruptcy Model-Averaged Estimates for Transportation Firms

In Table 8, "Post. Mean/(SE)" reports the model-averaged mean parameter estimate with the averaged standard error (reported in parentheses). "PIP" is the posterior inclusion probability of each variable. Each reported result is the average of estimates from $2^{19} = 524{,}288$ hazard models with the model prior, as noted in the last row. The uniform prior assigns equal weight to all models. The dilution prior weights each model by the determinant of the cross-correlation matrix of the included variables. The horizon of bankruptcy prediction is 12 months. All coefficients are hazard model coefficients. *N* is the number of firm-months in the sample. Variable definitions are provided in Table 1.

	$N = 51{,}189$					
	1		2		3	
Variable	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP
WC/TA	−0.0043 (0.0488)	0.0374	−0.0061 (0.0574)	0.0500	−0.0020 (0.0332)	0.0181
RE/TA	−0.0003 (0.0075)	0.0350	−0.0002 (0.0083)	0.0432	−0.0002 (0.0053)	0.0175
EBIT/TA	−0.0145 (0.0800)	0.0498	−0.0162 (0.0841)	0.0566	−0.0081 (0.0600)	0.0271
ME/TL	0.0002 (0.0015)	0.0449	0.0002 (0.0016)	0.0544	0.0001 (0.0010)	0.0206
S/TA	−0.0510 (0.1534)	0.1436	−0.0425 (0.1404)	0.1269	−0.0219 (0.1031)	0.0618
TL/TA	1.0433 (0.1728)	1.0000	1.0253 (0.1765)	1.0000	1.0348 (0.1638)	1.0000
CA/CL	−0.0001 (0.0043)	0.0392	−0.0001 (0.0048)	0.0473	0.0000 (0.0027)	0.0172
NI/TA	−0.0555 (0.1668)	0.1364	−0.0541 (0.1638)	0.1375	−0.0326 (0.1310)	0.0780
π_{MERTON}	0.0106 (0.0552)	0.0667	0.0161 (0.0682)	0.0885	0.0043 (0.0345)	0.0285
SIGMA	0.0932 (0.2317)	0.1813	0.1046 (0.2405)	0.2127	0.0696 (0.2045)	0.1327
AGE	−0.0162 (0.0455)	0.1526	−0.0149 (0.0437)	0.1448	−0.0066 (0.0300)	0.0635
RSIZE	0.0381 (0.0631)	0.3294	0.0711 (0.0755)	0.5573	0.0204 (0.0486)	0.1879
EXRET	−0.0141 (0.0732)	0.0716	−0.0118 (0.0675)	0.0667	−0.0070 (0.0516)	0.0318
CASH/MTA	0.0243 (0.1148)	0.0784	0.0274 (0.1222)	0.0885	0.0116 (0.0792)	0.0366
MB	0.000282 (0.00350)	0.0384	0.000258 (0.00381)	0.0483	0.000135 (0.00237)	0.0185
PRICE	−0.1340 (0.1063)	0.7186	−0.1777 (0.1067)	0.8356	−0.0920 (0.1012)	0.5347
$1/\sigma_E$	−0.5823 (0.1086)	0.9999	−0.5682 (0.1097)	0.9999	−0.6016 (0.1048)	0.9999
ME	−0.0016 (0.0040)	0.2027	−0.0034 (0.0054)	0.3745	−0.0008 (0.0029)	0.0975
F	0.0001 (0.0010)	0.0437	0.0001 (0.0011)	0.0513	0.0000 (0.0006)	0.0187
Bankruptcies	111		111		111	
Prior	Dilution		Uniform		Expected size = 5	

the data provide evidence against a variable if its posterior inclusion probability is below 0.1.[16]

[16]The choice of 0.1 and 0.9 as critical values is based on an equivalence between posterior inclusion probabilities and values of Bayes factors between models. See Jeffreys (1961) and Raftery (1995) for formal discussion and derivation of this result.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

Table 8 shows that among transportation, communications, and utilities firms, the only variables with a high posterior inclusion probability under any prior are TL/TA and $1/\sigma_E$. In contrast, the data suggest excluding WC/TA, RE/TA, EBIT/TA, ME/TL, CA/CL, π_{MERTON}, EXRET, CASH/MTA, MB, and F under the dilution prior. For the remaining variables, their middling posterior inclusion probabilities show that the data do not allow us to draw strong conclusions as to their importance.

The columns of Table 8 show the effects of the dilution prior. Under the dilution prior in column 1, nearly all of the variables in the highly correlated groups mentioned previously (RE/TA, EBIT/TA, and NI/TA; WC/TA, ME/TL, TL/TA, CA/CL, and CASH/MTA; RSIZE, PRICE, $1/\sigma_E$, and ME) have lower posterior inclusion probabilities than under the uniform prior in column 2. This effect is especially strong for RSIZE, PRICE, and ME, as the variables lose 23, 12, and 17 percentage points of posterior inclusion probability, respectively.

To see how the dilution prior helps determine which of a set of correlated variables is most effective in predicting bankruptcy, note that RSIZE, PRICE, and ME all have drops in posterior inclusion probability under the dilution prior, whereas $1/\sigma_E$ does not, despite high correlations between these three variables. The dilution prior puts less weight on models containing combinations of the four variables, increasing the relative importance of models containing only one of these variables. This shows that the good fit of models containing these variables results from the inclusion of $1/\sigma_E$, whereas the other variables add little to the model fit. Under the uniform prior, the other covariates receive relatively more credit for the good fit of models that also include $1/\sigma_E$, boosting their posterior inclusion probabilities.

Using the prior for an expected model size of five covariates, as in column 3 of Table 8, also lowers the posterior inclusion probabilities of a number of variables relative to the uniform prior but does so mechanically by lowering the prior inclusion probability for every variable. The difference in the effect of the smaller expected model size prior and the dilution prior can be seen in the posterior inclusion probabilities of such variables as S/TA, AGE, and EXRET. These three variables have lower posterior inclusion probabilities under the expected model size prior than under the uniform prior and slightly higher inclusion probabilities under the dilution prior, because they are not strongly correlated with many of the other potential explanatory variables. Because the dilution prior is effective in distinguishing among highly correlated individual covariates and because changes in posterior inclusion probabilities from the uniform prior are interpretable as reflecting correlations with other variables rather than a prior preference for a smaller model, it is the preferred prior for determining which variables are correlated with bankruptcy or default after accounting for model uncertainty.

Table 9 reports estimates for bankruptcies in the other industry groups and the sample of all firms using the dilution prior. Column 1 shows that for manufacturing firms, only TL/TA and $1/\sigma_E$ have high posterior inclusion probabilities, and the data recommend excluding 9 of the 19 variables. Column 2 shows the results for retail firms, the industry with the highest bankruptcy rate in the sample. TL/TA and $1/\sigma_E$ emerge as robust correlates of bankruptcy, and 11 variables fall below the cutoff for exclusion. Column 3 reveals that for service firms, S/TA,

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 9

Bankruptcy Model-Averaged Estimates by Industry

In Table 9, "Post. Mean/(SE)" reports the model-averaged mean parameter estimate with the averaged standard error (reported in parentheses). "PIP" is the posterior inclusion probability of each variable. Each reported result is the average of estimates from $2^{19} = 524{,}288$ hazard models with the model prior as noted in the last row. The dilution prior weights each model by the determinant of the cross-correlation matrix of the included variables. The horizon of bankruptcy prediction is 12 months. All coefficients are hazard model coefficients. *N* is the number of firm-months in the sample. Variable definitions are provided in Table 1.

	Industry							
	Manufacturing		Retail		Service		All Firms	
	1		2		3		4	
Variable	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP
WC/TA	0.0007 (0.0165)	0.0170	0.0235 (0.1106)	0.0855	−0.0034 (0.0480)	0.0241	0.0073 (0.0436)	0.0390
RE/TA	0.0009 (0.0053)	0.0452	0.1126 (0.1576)	0.4619	−0.0011 (0.0086)	0.0342	0.0430 (0.0264)	0.8160
EBIT/TA	−0.1942 (0.2958)	0.3487	−0.0030 (0.0375)	0.0151	−0.2028 (0.4557)	0.2007	−0.1662 (0.3232)	0.3965
ME/TL	−0.0132 (0.0555)	0.0947	−0.0148 (0.0271)	0.3080	0.00003 (0.0009)	0.0316	−0.0135 (0.0436)	0.0987
S/TA	−0.0109 (0.0612)	0.0490	−0.5522 (0.3782)	0.8244	−1.3605 (0.3842)	0.9925	−0.0099 (0.0544)	0.0395
TL/TA	0.4398 (0.1644)	0.9794	1.2597 (0.4971)	0.9964	1.2929 (0.2411)	0.9993	1.1083 (0.1318)	1.0000
CA/CL	−0.0025 (0.0097)	0.0810	0.0008 (0.0082)	0.0471	−0.0001 (0.0043)	0.0276	0.0000 (0.0019)	0.0103
NI/TA	−0.2476 (0.2892)	0.4826	−0.0091 (0.0687)	0.0330	−0.5019 (0.6226)	0.4589	−0.0073 (0.0895)	0.0069
π_{MERTON}	0.0489 (0.0927)	0.4051	0.0098 (0.0632)	0.0535	0.0926 (0.2256)	0.2987	0.0779 (0.0891)	0.6012
SIGMA	−0.0007 (0.0257)	0.0171	0.0097 (0.0831)	0.0327	0.0088 (0.1084)	0.0307	0.0024 (0.0342)	0.0133
AGE	−0.0019 (0.0103)	0.0534	−0.0032 (0.0235)	0.0601	−0.1216 (0.1232)	0.5650	−0.0797 (0.0578)	0.7125
RSIZE	0.0107 (0.0339)	0.2574	0.0002 (0.0226)	0.0583	0.0317 (0.0603)	0.2518	0.0167 (0.0275)	0.4561
EXRET	−0.2423 (0.1495)	0.8164	−0.0920 (0.2217)	0.2066	−0.1857 (0.2886)	0.3505	−0.0545 (0.1119)	0.3996
CASH/MTA	−0.4049 (0.2398)	0.8296	−0.0067 (0.0784)	0.0297	−0.2266 (0.4115)	0.2855	−0.0734 (0.2196)	0.2643
MB	0.00166 (0.00624)	0.0915	−0.00234 (0.0146)	0.0664	0.000698 (0.00695)	0.0375	−0.00135 (0.00607)	0.0633
PRICE	−0.0188 (0.0317)	0.4391	−0.1386 (0.1508)	0.5664	−0.3214 (0.2181)	0.7623	−0.0324 (0.0364)	0.6201
$1/\sigma_E$	−0.2179 (0.0599)	0.9988	−0.4597 (0.1874)	0.9830	−0.3104 (0.1163)	0.9800	−0.4928 (0.0594)	1.0000
ME	−0.0370 (0.0248)	0.8779	−0.0158 (0.0184)	0.5384	0.00002 (0.0008)	0.0381	−0.0145 (0.0470)	0.0995
F	0.0001 (0.0010)	0.0259	−0.0035 (0.0173)	0.0831	0.0020 (0.0078)	0.0877	0.0000 (0.0006)	0.0133
N	257,576		44,435		133,338		544,422	
Bankruptcies	275		95		156		721	
Prior	Dilution		Dilution		Dilution		Dilution	

TL/TA, and $1/\sigma_E$ are the only variables whose inclusion is strongly supported by the data, and eight variables are recommended for exclusion. Column 4 imposes the restriction that the probability of bankruptcy for all firms responds in the same way to changes in each variable and estimates over the sample of all firms. This restriction is rather weak; only TL/TA and $1/\sigma_E$ have high posterior inclusion

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

probabilities in the larger sample, and nine variables are recommended for exclusion. Throughout Tables 8 and 9, posterior inclusion probabilities and t-tests select the same variables as robust correlates at conventional significance levels.

These results show that only TL/TA and $1/\sigma_E$ are robust correlates of bankruptcy in every industry group and the overall sample. TL/TA is an accounting proxy for a firm's indebtedness relative to its assets or income, as evidenced by high correlations with WC/TA, ME/TL, CA/CL, and CASH/MTA. Similarly, $1/\sigma_E$ is a measure of the volatility of the firm's market equity and is highly correlated with SIGMA, RSIZE, and PRICE. Because these two variables are correlated with bankruptcy even after considering model uncertainty in all industry groups and using dilution priors to control for their correlations with other variables, this is strong evidence that TL/TA and $1/\sigma_E$ are the best available predictors and should be included in all firm bankruptcy studies.

The results show some evidence of parameter differences across industry groups, as S/TA is a robust predictor of bankruptcy for service firms. However, the similarities in robust bankruptcy predictors across industry groups indicate that model uncertainty is a greater source of variability in parameter estimates than exchangeability uncertainty, where observations in different industry groups are generated by different underlying statistical models.[17] Firms in different industries may face different market conditions, competitive pressures, or industry-specific shocks such that firms with identical financial indicators in different industries have different probabilities of filing for bankruptcy, but cross-industry parameter variation appears to have less influence on which variables appear significant than model selection. Although these industry groups mirror those used by Chava and Jarrow (2004), it is possible that the choice of groups drives the result shown here. However, these industry groups have varying numbers of observations, with some samples much smaller than the overall sample of firms, and differences in observables, as described in Tables 4 and 6. This suggests that the uniformity of the importance of TL/TA and $1/\sigma_E$ is likely not a function of sample size or choice of industry grouping. I quantify the relative importance of model uncertainty and parameter variation across industry groups in the out-of-sample forecast results.

The data consistently recommend that several variables not be used as predictors. In the overall sample, the data reject WC/TA, ME/TL, S/TA, CA/CL, NI/TA, SIGMA, MB, ME, and F, although S/TA still appears to be a good predictor for service firms. Many variables are also rejected in the industry subsamples: The data reject WC/TA, CA/CL, MB, and F in all four subsamples and ME/TL and SIGMA in three subsamples, indicating that these variables provide little help even in predictions for specific industries. Table 2 indicates that these variables have been used in many studies, and the appearance of both market and accounting variables on this list shows that the model-averaging procedure does not select variables based on the frequency of data availability. The high correlation of some of these variables with TL/TA and $1/\sigma_E$ combined with the dilution prior also does not explain this finding, as this correlation equally and symmetrically affects the very high posterior inclusion probabilities of TL/TA and $1/\sigma_E$. Thus, the data

[17] See Brock and Durlauf (2001) and Durlauf, Johnson, and Temple (2005) for formal discussions of how this form of uncertainty is related to the exchangeability of random variables.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

show that there is little empirical support for including these variables in a firm bankruptcy model, especially those rejected across industry groups that appear to be of minimal use even in studies of specific industries.

Table 10 reveals that TL/TA and $1/\sigma_E$ are also the best available predictors of defaults. Column 1 shows that across all firms, only TL/TA and $1/\sigma_E$ are

TABLE 10
Default Model-Averaged Estimates by Industry

In Table 10, "Post. Mean/(SE)" reports the model-averaged mean parameter estimate with the averaged standard error (reported in parentheses). "PIP" is the posterior inclusion probability of each variable. Each reported result is the average of estimates from $2^{19}=524{,}288$ hazard models with the model prior as noted in the last row. The dilution prior weights each model by the determinant of the cross-correlation matrix of the included variables. The horizon of default prediction is 12 months. All coefficients are hazard model coefficients. *N* is the number of firm-months in the sample. Variable definitions are provided in Table 1.

	Industry									
	All Firms		Manufacturing		Transportation		Retail		Service	
	1		2		3		4		5	
Variable	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP	Post. Mean/ (SE)	PIP
WC/TA	0.0033 (0.0553)	0.0567	0.0018 (0.0543)	0.0196	0.0079 (0.0723)	0.0845	0.0100 (0.1494)	0.0544	−0.0021 (0.0589)	0.0316
RE/TA	0.0253 (0.0650)	0.3029	0.0524 (0.1790)	0.2222	0.0005 (0.0100)	0.0451	0.1100 (0.1832)	0.4440	0.0048 (0.0214)	0.1791
EBIT/TA	−0.1477 (0.2956)	0.3797	−0.1680 (0.2676)	0.4699	−0.0128 (0.0735)	0.1398	−0.0028 (0.0455)	0.0461	−0.1767 (0.4208)	0.3306
ME/TL	−0.0060 (0.0563)	0.0857	−0.0029 (0.0480)	0.0447	0.0001 (0.0018)	0.0487	−0.0044 (0.0367)	0.0946	−0.0005 (0.0048)	0.0826
S/TA	−0.0137 (0.1871)	0.0613	−0.0331 (0.1058)	0.2387	−0.2619 (0.7208)	0.2894	−0.0480 (0.3703)	0.1112	−1.2125 (0.4136)	0.9910
TL/TA	1.0814 (0.1374)	1.0000	0.7237 (0.2019)	0.9997	1.0760 (0.1789)	1.0000	1.3488 (0.6633)	0.9580	1.4081 (0.3761)	0.9998
CA/CL	0.0001 (0.0020)	0.0391	−0.0011 (0.0104)	0.0813	−0.0005 (0.0045)	0.0933	0.0004 (0.0150)	0.0149	0.0005 (0.0065)	0.0573
NI/TA	−0.0031 (0.0953)	0.0303	−0.0620 (0.2717)	0.1779	−0.0485 (0.1508)	0.2549	−0.0083 (0.0774)	0.0947	−0.1453 (0.5697)	0.2018
π_{MERTON}	0.1099 (0.1590)	0.5091	0.1916 (0.3486)	0.4171	0.0104 (0.0554)	0.1576	0.0184 (0.1070)	0.1413	0.1338 (0.2703)	0.3857
SIGMA	0.0018 (0.0418)	0.0409	−0.0003 (0.0413)	0.0023	0.0847 (0.2192)	0.3053	0.0078 (0.1053)	0.0571	0.0055 (0.1294)	0.0360
AGE	−0.0700 (0.0525)	0.8174	−0.0015 (0.0118)	0.1075	−0.0154 (0.0446)	0.2708	−0.0021 (0.0352)	0.0402	−0.1091 (0.1185)	0.6496
RSIZE	0.0276 (0.0583)	0.3723	0.0133 (0.0392)	0.2732	0.0347 (0.0600)	0.4276	0.0019 (0.0276)	0.0645	0.0357 (0.0651)	0.4135
EXRET	−0.0653 (0.1140)	0.4346	−0.1283 (0.1890)	0.4937	−0.0163 (0.0787)	0.1627	−0.0755 (0.2461)	0.2341	−0.1809 (0.3060)	0.4506
CASH/MTA	−0.0638 (0.1945)	0.2650	−0.1177 (0.2331)	0.3919	0.0413 (0.1390)	0.2237	−0.0092 (0.1082)	0.0758	−0.1981 (0.3928)	0.3773
MB	−0.0006 (0.0054)	0.0855	0.0006 (0.0060)	0.0813	0.0003 (0.0031)	0.0856	−0.0015 (0.0132)	0.0951	0.0007 (0.0063)	0.0827
PRICE	−0.0400 (0.0371)	0.7123	−0.0458 (0.0458)	0.6735	−0.1175 (0.0986)	0.7628	−0.1196 (0.1432)	0.5923	−0.2784 (0.1960)	0.8364
$1/\sigma_E$	−0.4479 (0.0562)	1.0000	−0.1912 (0.0611)	0.9982	−0.5799 (0.1101)	1.0000	−0.4101 (0.1960)	0.9567	−0.2685 (0.1012)	0.9961
ME	−0.0021 (0.0409)	0.0383	−0.0324 (0.0227)	0.8552	−0.0002 (0.0037)	0.0545	−0.0017 (0.0194)	0.0713	0.0001 (0.0009)	0.0964
F	−0.0001 (0.0008)	0.0832	0.0001 (0.0013)	0.0445	0.0001 (0.0011)	0.0372	−0.0018 (0.0191)	0.0680	0.0010 (0.0096)	0.0718
N	544,422		257,576		51,189		44,435		133,338	
Defaults	834		305		153		101		174	
Prior	Dilution		Dilution		Dilution		Dilution		Dilution	

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

robustly correlated with default, and the data recommend excluding nine variables. Column 2 reports that for manufacturing firms, TL/TA and $1/\sigma_E$ have high posterior inclusion probabilities, whereas six variables are below the threshold for exclusion. Column 3 contains results for transportation firms, indicating that both TL/TA and $1/\sigma_E$ are supported by the data, whereas seven variables are recommended for exclusion. Column 4 shows TL/TA and $1/\sigma_E$ as the only robust correlates of default for retail firms, with 12 variables below the exclusion threshold. The posterior inclusion probabilities for TL/TA and $1/\sigma_E$ are the lowest in this subsample of any industry group, perhaps reflecting the smaller number of defaults in this industry group despite the high default rate. Column 5 reports the results for service firms, with TL/TA, $1/\sigma_E$, and S/TA being robust correlates of default, while the data recommend excluding seven variables.

The data recommend excluding several variables from default prediction models. WC/TA, ME/TL, S/TA, CA/CL, NI/TA, SIGMA, MB, ME, and F are rejected in the overall sample, with WC/TA, ME/TL, CA/CL, MB, and F rejected in every industry group and SIGMA and ME rejected in three of the four groups. These variables are broadly consistent with those rejected in the bankruptcy-only sample, although more variables appear to be rejected consistently across industry groups. With the notable exception of S/TA for service firms, these variables receive little support from the data across industry groups and are unlikely to be useful in default prediction.

For default prediction, the data strongly support using TL/TA and $1/\sigma_E$ both in the overall sample and in the industry-group samples. As before, S/TA emerges as a robust default correlate for service firms, but low posterior inclusion probabilities for other industry groups lead to its rejection in the overall sample. Overall, only TL/TA and $1/\sigma_E$ consistently have high posterior inclusion probabilities for default prediction, indicating that they are the most robust correlates of default. I test the potential significance of the variation in parameter values across industry groups in the forecasting results.

B. Kitchen Sink Regressions

To determine the practical magnitude of model uncertainty, I run "kitchen sink" regressions in which all available covariates are explanatory variables in a single hazard model. The kitchen sink model is a natural comparison for the model-averaging results, because one objection to the use of model averaging is that a model with all available covariates allows all parameters to converge to their true values as the amount of data increases.[18] Because the kitchen sink results do not take into account model uncertainty, the kitchen sink regression is equivalent to performing model averaging with a prior probability of 1 on the model with all covariates, and 0 on all other models. Table 11 shows the estimates from the kitchen sink regressions predicting default at a horizon of 12 months for all firms and all four industries.[19]

[18] See Sala-i-Martin et al. (2004) and Durlauf et al. (2008) for discussion of kitchen sink estimates and model averaging in the context of linear models.

[19] Results for models that predict bankruptcies only are similar and are presented in Table 12.

Comparing the results in column 1 of Table 11 with those in column 1 of Table 10 reveals that in the kitchen sink regression for all firms, 15 of the 19 variables are statistically significant at the 10% level or higher under conservative standard errors of the form suggested by Shumway (2001) and clustered at the firm level. In contrast, the model-averaged results indicate that only 2 of these 15 variables are strongly correlated with bankruptcy after accounting for model uncertainty, and 5 of the 15 should be excluded from the model. The large differences between the model-averaged and kitchen sink results show that evaluating

TABLE 11
Default Kitchen Sink Regressions

In Table 11, the column headers show the hazard models estimated for the dependent variables and industry groups. The horizon of all predictions is 12 months. All coefficients are hazard model coefficients. Standard errors (reported in parentheses) are Huber–White robust estimates, clustered at the firm level. *N* is the number of firm-months in the sample. Variable definitions are provided in Table 1. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Industry				
	All Firms	Manufacturing	Transportation	Retail	Service
Variable	1	2	3	4	5
WC/TA	0.1039***	0.0647**	−0.0161	0.0507	0.0166
	(0.0398)	(0.0319)	(0.1036)	(0.0564)	(0.0833)
RE/TA	0.0247***	0.0098**	0.0362*	0.0683***	0.0020
	(0.0060)	(0.0042)	(0.0186)	(0.0263)	(0.0104)
EBIT/TA	−0.5934***	−0.1969**	−1.0770***	−0.3637	−0.4163*
	(0.1374)	(0.0982)	(0.3696)	(0.2694)	(0.2364)
ME/TL	−0.0047**	−0.0035	0.0015	−0.0089	−0.0023
	(0.0020)	(0.0023)	(0.0029)	(0.0084)	(0.0018)
S/TA	−0.1025***	−0.0619*	−0.1256	−0.1052**	−0.2936***
	(0.0364)	(0.0375)	(0.0875)	(0.0513)	(0.0908)
TL/TA	0.4132***	0.1697***	0.4603***	0.2447***	0.2620***
	(0.0404)	(0.0348)	(0.0791)	(0.0753)	(0.0801)
CA/CL	−0.0035	−0.0054	−0.0091	0.0047	0.0055
	(0.0042)	(0.0039)	(0.0101)	(0.0053)	(0.0069)
NI/TA	−0.0840	−0.0990	0.1383	−0.2092	−0.1036
	(0.1021)	(0.0675)	(0.2699)	(0.2144)	(0.1720)
π_{MERTON}	0.2147***	0.1359***	0.0585	0.0254	0.2615***
	(0.0248)	(0.0183)	(0.0551)	(0.0387)	(0.0577)
SIGMA	0.0420	−0.0451	0.3118**	0.1117	−0.0482
	(0.0683)	(0.0562)	(0.1312)	(0.1133)	(0.1528)
AGE	−0.0300***	−0.0076	−0.0105	−0.0148	−0.0576***
	(0.0093)	(0.0074)	(0.0222)	(0.0132)	(0.0194)
RSIZE	0.0524***	0.0329***	0.0597***	0.0391***	0.0398**
	(0.0088)	(0.0082)	(0.0167)	(0.0134)	(0.0154)
EXRET	−0.1434***	−0.0725***	−0.0292	−0.0530	−0.1272**
	(0.0329)	(0.0273)	(0.0647)	(0.0436)	(0.0632)
CASH/MTA	−0.3012***	−0.1648***	0.1391	−0.1468	−0.3323***
	(0.0680)	(0.0560)	(0.1130)	(0.1135)	(0.1238)
MB	−0.0010***	−0.0007**	−0.0008	−0.0004	0.0003
	(0.0004)	(0.0003)	(0.0005)	(0.0008)	(0.0008)
PRICE	−0.0939***	−0.0559***	−0.0822***	−0.0664***	−0.0890***
	(0.0099)	(0.0080)	(0.0208)	(0.0158)	(0.0231)
$1/\sigma_E$	−0.1293***	−0.0572***	−0.1469***	−0.0648***	−0.0873**
	(0.0164)	(0.0126)	(0.0339)	(0.0243)	(0.0437)
ME	−0.0040**	−0.0088*	−0.0025**	−0.0084	−0.0014
	(0.0020)	(0.0046)	(0.0011)	(0.0060)	(0.0015)
F	0.0009	0.0002	0.0004	−0.0065	0.0064
	(0.0011)	(0.0009)	(0.0015)	(0.0069)	(0.0044)
N	544,422	257,576	51,189	44,435	133,338

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 12
Bankruptcy Kitchen Sink Regressions

In Table 12, the column headers show the hazard models estimated for the dependent variables and industry groups. The horizon of all predictions is 12 months. All coefficients are hazard model coefficients. Standard errors (reported in parentheses) are Huber–White robust estimates, clustered at the firm level. *N* is the number of firm-months in the sample. Variable definitions are provided in Table 1. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Industry				
	All Firms	Manufacturing	Transportation	Retail	Service
Variable	1	2	3	4	5
WC/TA	0.1014**	0.0646*	−0.0630	0.0506	0.0322
	(0.0409)	(0.0338)	(0.111)	(0.0585)	(0.0805)
RE/TA	0.0229***	0.0100**	0.0325*	0.0685**	0.0016
	(0.0061)	(0.0044)	(0.019)	(0.0273)	(0.01)
EBIT/TA	−0.6059***	−0.2149**	−1.0061***	−0.3768	−0.3863*
	(0.1427)	(0.1022)	(0.3843)	(0.2888)	(0.2327)
ME/TL	−0.0043**	−0.0033	0.0017	−0.0094	−0.0022
	(0.002)	(0.0023)	(0.0028)	(0.0093)	(0.0018)
S/TA	−0.0690*	−0.0489	−0.0312	−0.1059**	−0.2960***
	(0.0363)	(0.0376)	(0.0828)	(0.0512)	(0.0847)
TL/TA	0.3748***	0.1575***	0.3678***	0.2334***	0.2322***
	(0.0413)	(0.0363)	(0.08)	(0.0777)	(0.0759)
CA/CL	−0.0040	−0.0068	−0.0071	0.0045	0.0050
	(0.0045)	(0.0043)	(0.0108)	(0.0056)	(0.0066)
NI/TA	−0.1085	−0.1025	0.0241	−0.1908	−0.1323
	(0.1059)	(0.0726)	(0.2822)	(0.236)	(0.1684)
π_{MERTON}	0.2160***	0.1247***	0.0877	0.0285	0.2595***
	(0.0254)	(0.0187)	(0.059)	(0.0392)	(0.0553)
SIGMA	0.0529	−0.0217	0.2936**	0.1501	−0.0661
	(0.0707)	(0.0568)	(0.1425)	(0.1123)	(0.1524)
AGE	−0.0289***	−0.0054	−0.0110	−0.0146	−0.0481***
	(0.0095)	(0.0073)	(0.0238)	(0.0138)	(0.0186)
RSIZE	0.0477***	0.0305***	0.0574***	0.0361***	0.0343**
	(0.0095)	(0.0085)	(0.019)	(0.0137)	(0.0153)
EXRET	−0.1468***	−0.0735***	−0.0308	−0.0574	−0.1164*
	(0.0344)	(0.0284)	(0.0691)	(0.0461)	(0.0636)
CASH/MTA	−0.2839***	−0.1524***	0.1699	−0.1257	−0.3161***
	(0.0674)	(0.0557)	(0.1158)	(0.1084)	(0.1169)
MB	−0.0007*	−0.0005*	−0.0002	−0.0001	0.0003
	(0.0004)	(0.0003)	(0.0006)	(0.0007)	(0.0007)
PRICE	−0.0951***	−0.0559***	−0.0848***	−0.0648***	−0.0880***
	(0.0101)	(0.0083)	(0.0214)	(0.0163)	(0.0214)
$1/\sigma_E$	−0.1329***	−0.0603***	−0.1407***	−0.0633**	−0.1140**
	(0.0175)	(0.0133)	(0.0368)	(0.0249)	(0.0473)
ME	−0.0040*	−0.0087*	−0.0029*	−0.0085	−0.0007
	(0.0024)	(0.0047)	(0.0016)	(0.007)	(0.0014)
F	0.0004	0.0007	0.0003	−0.0068	0.0040
	(0.0013)	(0.0009)	(0.0018)	(0.0076)	(0.0043)
N	544,422	257,576	51,189	44,435	133,338

a variable by its statistical significance in a kitchen sink regression and ignoring model uncertainty overstates the strength of the relationship between the variable and bankruptcy.

Results for specific industries show a similar pattern. Comparing column 2 of Table 11 with column 2 of Table 10, the kitchen sink regression finds 13 variables to be significant predictors of default in the manufacturing sector, whereas model averaging selects only 2 of these as robust to model uncertainty. Column 3 of Table 11 and column 3 of Table 10 show that for the transportation, communications, and utilities industry group, the kitchen sink regression finds eight

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

significant predictors compared with two from model averaging. In the retail industry, column 4 of Table 11 indicates six statistically significant variables in the kitchen sink regression, whereas column 4 of Table 10 reports only two robust correlates. Finally, column 5 of Table 11 shows that for service firms, 10 variables are statistically significant at conventional levels, whereas column 5 of Table 10 shows only 3 of these to be robust to model uncertainty. In every case, the variables robust to model uncertainty are a subset of those statistically significant in the kitchen sink regression.

Taken together, the kitchen sink regressions show that simply including all available covariates leads to claims of statistical significance that are not robust to model uncertainty. Failing to consider this source of uncertainty creates overconfidence in determining which variables are good predictors of default, and the large difference in the number of significant predictors between the kitchen sink regression and the model-averaging results shows that this overconfidence is empirically relevant in magnitude. In contrast, Bayesian model averaging accounts for model uncertainty by estimating many models, explicitly incorporating between-model variance and mitigating the effects of correlations between explanatory variables by including parameter estimates from models with less multicollinearity. As a result, model averaging selects far fewer variables as correlated with impending firm default than the number found to be significant in the kitchen sink regression.

V. Out-of-Sample Forecasting

I now evaluate the ability of Bayesian model averaging to produce accuracy gains in out-of-sample forecasts. I first use data over the period 1987–2000 to estimate each of the 2^{19} possible models. With these coefficients, I predict the probability under each model that a firm will default in 12 months for every firm-month over the out-of-sample period 2001–2008. The model-averaged forecast for a given firm-month is the weighted average of the 2^{19} different forecasts for that firm-month, where each forecast is weighted by the posterior probability of the model that generated it. I present results for all defaults; the results for only bankruptcies are similar.

I compare the model-averaged forecasts with the forecasts of the kitchen sink model over the period 1987–2000. I also estimate the kitchen sink model with random effects at the firm level, a standard method of allowing for unobserved firm-level heterogeneity. I compute the forecasts implied by the kitchen sink model and the kitchen sink model with random effects for every firm-month from 2001 to 2008 to compare against model-averaged forecasts using the dilution prior.[20] I also use the model proposed by Campbell et al. (2008), denoted CHS, and the best 5-variable model that can be formed from the given 19 variables according to the BIC criterion, denoted BIC5, to test the model-averaging forecast against other popular, parsimonious forecasting models and model-selection methods.[21]

[20] Results using the other two priors described previously, not reported here (available from the author), are similar.

[21] For bankruptcy prediction, BIC5 includes TL/TA, π_{MERTON}, EXRET, PRICE, and $1/\sigma_E$. For default prediction, BIC5 includes TL/TA, π_{MERTON}, RSIZE, PRICE, and $1/\sigma_E$. Campbell et al. (2008) do not specifically forecast defaults, but they do forecast business failures that include similar events.

To compare the relative accuracies of each forecast, I score the forecasts using the predictive logarithmic scoring rule

$$\text{PLS} = \sum_i \sum_t \left(Y_{it} \times \ln(p_{it}) + (1 - Y_{it}) \times \ln(1 - p_{it})\right),$$

where p_{it} is the predicted probability that firm i will default 12 months after month t, and Y_{it} is a dummy variable equal to 1 if firm i defaults 12 months after month t. A higher predictive log score indicates a more accurate forecast. The difference in predictive accuracy per firm is given in percentage terms by

$$\text{PLS}_{\text{diff}} = \left[\exp\left(\frac{\text{PLS}_{\text{BMA}} - \text{PLS}_{\text{REG}}}{n_f}\right) - 1\right] \times 100, \tag{6}$$

where PLS_{BMA} and PLS_{REG} are the predictive log scores from the model-averaged and the regression forecasts, respectively, and n_f is the number of firms in the prediction sample. Predictive log scoring is appropriate in bankruptcy and default forecasting, because filing and defaulting are binary outcomes.[22] This criterion is demanding, because each of the predicted monthly probabilities is small, so even small changes in the predicted probability of a default in a given month can have relatively large effects on the predicted log score. However, the results of Campbell et al. (2008) show that firms separated by only 1% in the distribution of predicted default probabilities can have very different stock returns. Thus, small changes in predicted probabilities of business failure, especially at the very top and bottom of the empirical probability distribution, can have large effects on other firm outcomes. In this light, the sensitivity of the predictive log score to small changes in the predicted probability of bankruptcy or default is desirable.

I also calculate the area under the receiver operating characteristic (ROC) curve and the percentage of defaults occurring in each decile of the distribution of predicted values, criteria used by Shumway (2001), Chava and Jarrow (2004), Bharath and Shumway (2008), and Giordani et al. (2014). By relying on the rank order of firm-months, the decile method is less influenced by small changes in the predicted probabilities of failure, because they are unlikely to change the decile in which a firm-month lies in the distribution. These measures of forecast accuracy complement the predictive log score by showing approximately where in the distribution of predicted default probabilities the model-averaged and other regression forecasts differ and by giving a more familiar sense of scale to the percentage differences found using predictive log scoring.

Table 13 shows the difference in default forecast accuracy between the model-averaged forecasts with the dilution prior and other forecasting models.[23] Focusing on Panel A, Bayesian model averaging yields an improvement in forecast accuracy relative to the kitchen sink approach in the sample of all firms of nearly 4% per firm. Forecast accuracy gains are also large within industry groups; model-averaged forecasts produce accuracy gains of approximately 5% per firm in the manufacturing industry, 1.2% per firm in the transportation industry, 2% per

[22] Winkler (1969) and Schervish (1989) discuss scoring rules for forecasting binary outcomes. Using root mean squared predictive error produces qualitatively similar results.

[23] Table 14 presents analogous results for bankruptcies.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 13
Default Forecast Accuracy Gains

In Table 13, out-of-sample predictions are from the period 2001–2008, while predictions are generated using data from 1987–2000. Panel A shows the difference for the industry in the predictive log score of the forecasts made by Bayesian model averaging using the dilution prior versus the forecasts made by the regression as given by equation (6). "Random Effects" is the kitchen sink model including firm-level random effects. "CHS" is the model of Campbell et al. (2008), with coefficients estimated on this sample. "BIC5" is the regression with five variables that best fits the 1987–2000 data according to BIC. Panel B compares the predictive log score from model-averaged forecasts allowing coefficients to differ by industry to the predictive log score from model-averaged forecasts where coefficients are constrained to be the same across industries. Panel C compares the predictive log score for the model with variables to the predictive log score of the kitchen sink model forecast.

Panel A. Dilution Prior versus Baseline Regressions

	Industry Group				
Regression	All Firms	Manufacturing	Transportation	Retail	Service
Kitchen sink	3.76%	5.12%	1.20%	1.94%	1.40%
Random effects	0.11%	1.07%	0.30%	0.16%	4.03%
CHS	2.74%	2.13%	1.96%	1.33%	1.09%
BIC5	0.04%	0.99%	0.15%	0.36%	0.70%

Panel B. Industry-Specific versus Nonindustry-Specific Models

Prior	Manufacturing, Transportation, Retail, and Service Firms
Dilution	1.15%
Uniform	1.10%
Expected size = 5	1.11%

Panel C. Single Regressions versus Kitchen Sink Regression

	Industry Group				
Variables	All Firms	Manufacturing	Transportation	Retail	Service
TL/TA, $1/\sigma_E$	0.36%	−0.20%	0.49%	−0.12%	0.45%
Other 17 variables	−0.21%	−0.81%	0.23%	−0.20%	−0.48%

firm in the retail industry, and 1.4% per firm in the service industry. Comparing the model-averaged forecasts to the kitchen sink model with random effects shows that model averaging produces superior forecasts, although the relative improvement is smaller in all groups except the service industry. The model-averaged forecasts are also more accurate than the forecasts from both the CHS and BIC5 models, although by smaller margins against the BIC5 model.

In Panel B of Table 13, I compare the predictive log scores of the model-averaged forecasts using industry-specific coefficients and the model-averaged forecast where coefficients must be the same for all industries. The results show that using industry-specific coefficients improves out-of-sample forecast accuracy by approximately 1.1% per firm over using coefficients fixed across industry groups. Out-of-sample forecast accuracy improves when allowing for exchangeability uncertainty, just as it improves when using model-averaged forecasts. This result is in contrast to Chava and Jarrow (2004), who find no improvement in forecast accuracy when allowing coefficients to differ across industries. However, the improvement is less than one-third of the magnitude of the increase in forecast accuracy from using model averaging across all firms, as shown in the top panel. Accounting for parameter uncertainty across models appears to be more than three times as important in forecasting as accounting for parameter uncertainty across these industry groups.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 14
Bankruptcy Forecast Accuracy Gains

In Table 14, out-of-sample predictions are from the period 2001–2008, while predictions are generated using data from 1987–2000. Panel A shows the difference for the industry in the predictive log score of the forecasts made by Bayesian model averaging using the dilution prior versus the forecasts made by the regression as given by equation (6). "Random Effects" is the kitchen sink model including firm-level random effects. "CHS" is the model of Campbell et al. (2008), with coefficients estimated on this sample. "BIC5" is the regression with five variables that best fits the 1987–2000 data according to BIC. Panel B compares the predictive log score from model-averaged forecasts allowing coefficients to differ by industry to the predictive log score from model-averaged forecasts where coefficients are constrained to be the same across industries. Panel C compares the predictive log score for the model with variables to the predictive log score of the kitchen sink model forecast.

Panel A. Dilution Prior versus Baseline Regressions

	Industry Group				
Regression	All Firms	Manufacturing	Transportation	Retail	Service
Kitchen sink	3.99%	5.55%	1.33%	2.19%	1.66%
Random effects	0.10%	1.13%	0.28%	0.18%	4.73%
CHS	2.57%	2.23%	1.56%	1.31%	1.25%
BIC5	0.03%	0.99%	0.18%	0.41%	0.85%

Panel B. Industry-Specific versus Nonindustry-Specific Models

Prior	Manufacturing, Transportation, Retail, and Service Firms
Dilution	1.25%
Uniform	1.19%
Expected size = 5	1.21%

Panel C. Single Regressions versus Kitchen Sink Regression

	Industry Group				
Variables	All Firms	Manufacturing	Transportation	Retail	Service
TL/TA, $1/\sigma_E$	0.38%	−0.30%	0.57%	−0.15%	0.50%
Other 17 variables	−0.22%	−0.94%	0.28%	−0.23%	−0.55%

Finally, I compare forecasts from two selected models against the baseline kitchen sink regression in Panel C of Table 13. The model that includes only TL/TA and $1/\sigma_E$, the two variables shown to be most important when using model averaging on the full data, outperforms the kitchen sink model in the sample of all firms and in two of the four industry groups, suggesting that these variables have considerable predictive power even when used alone. By contrast, the model with the 17 other covariates that are not TL/TA or $1/\sigma_E$ shows a general decrease in predictive performance relative to the kitchen sink model in the sample of all firms and three of the four subsamples. In each case, the relative performance of the model with TL/TA and $1/\sigma_E$ is better than the model using all other covariates. Because the kitchen sink regression contains all variables, these results show that adding the other 17 variables to a model with only TL/TA and $1/\sigma_E$ as predictors generally does not result in better forecasts, whereas adding TL/TA and $1/\sigma_E$ to a model containing the other 17 variables generally does improve forecast accuracy.

Table 15 reports the area under the ROC curve and the percentage of defaults in each sample appearing in each decile of the predicted distribution for defaults.[24] In every sample, predictions from the model-averaged estimates appear more accurate than those from the kitchen sink model, with or without random effects. The individual model using only TL/TA and $1/\sigma_E$ does not perform as well as the

[24] Table 16 presents analogous results for bankruptcies.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 15
Defaults by Prediction Decile

In Table 15, entries are the percentage of actual defaults in the industry group in the listed decile of the predicted probability distribution of default across out-of-sample firm-months. Percentages may not sum to 100% due to rounding. Predicted probability distributions use estimates from the model listed at the top of each panel. Out-of-sample predictions are from the period 2001–2008, while predictions are generated using data from 1987–2000. "ROC" is the area under the receiver operating characteristic curve.

Decile	Dilution	Kitchen Sink	Random Effects	TL/TA, $1/\sigma_E$	CHS	BIC5
Panel A. All Firms						
1	66.96%	61.50%	65.77%	62.79%	67.02%	66.37%
2	12.98%	16.07%	13.58%	14.77%	11.36%	12.60%
3	5.63%	5.46%	5.57%	6.93%	5.30%	6.44%
4	3.95%	4.22%	4.11%	4.11%	3.79%	2.59%
5	3.51%	3.67%	2.71%	3.84%	2.43%	4.22%
6–10	6.97%	9.08%	8.28%	7.57%	10.11%	7.79%
ROC	0.875	0.851	0.869	0.855	0.858	0.870
Panel B. Manufacturing						
1	74.69%	61.48%	70.13%	55.41%	67.04%	70.71%
2	11.32%	14.93%	15.51%	17.12%	18.76%	14.36%
3	5.34%	5.34%	7.07%	10.96%	7.28%	7.28%
4	2.88%	7.28%	1.51%	6.13%	2.88%	0.94%
5	0.58%	2.31%	0.00%	4.04%	0.58%	1.51%
6–10	5.19%	8.65%	5.77%	6.35%	3.46%	5.19%
ROC	0.905	0.862	0.897	0.852	0.872	0.898
Panel C. Transportation						
1	47.70%	35.46%	41.92%	41.92%	48.37%	52.89%
2	20.89%	30.14%	19.76%	17.90%	21.95%	19.63%
3	16.63%	14.44%	15.37%	14.44%	9.71%	12.71%
4	9.12%	9.12%	6.92%	8.32%	6.12%	2.20%
5	1.73%	5.66%	5.66%	7.05%	0.00%	1.73%
6–10	3.93%	5.19%	10.38%	10.38%	13.84%	10.85%
ROC	0.828	0.802	0.812	0.803	0.779	0.823
Panel D. Retail						
1	55.11%	32.66%	49.96%	34.35%	49.96%	51.73%
2	13.84%	15.53%	12.07%	13.84%	12.07%	24.05%
3	13.84%	10.38%	10.38%	17.30%	10.38%	3.46%
4	3.46%	3.46%	6.92%	6.92%	6.92%	13.84%
5	1.69%	12.07%	5.15%	10.30%	5.15%	0.00%
6–10	12.07%	25.91%	15.53%	17.30%	15.53%	6.92%
ROC	0.849	0.730	0.805	0.733	0.793	0.830
Panel E. Service						
1	65.87%	51.92%	44.71%	58.27%	62.12%	47.59%
2	12.51%	11.56%	12.99%	14.44%	18.28%	16.83%
3	5.77%	5.77%	7.21%	4.81%	1.44%	5.30%
4	1.44%	5.77%	5.77%	6.06%	1.44%	2.88%
5	2.88%	9.13%	4.81%	3.17%	4.33%	1.44%
6–10	11.53%	15.86%	24.51%	13.26%	12.40%	25.95%
ROC	0.842	0.775	0.717	0.800	0.817	0.743

model-averaged estimates, but it is more accurate than the kitchen sink model in the overall sample and the transportation and service groups, as suggested by the predictive log scores.

Across all firm groups, the gains in predictive accuracy from the model-averaging approach relative to the more standard approaches appear throughout the decile distribution. These improvements in decile predictions are similar in magnitude to those reported by Shumway ((2001), p. 122) when adding market

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

TABLE 16
Bankruptcies by Prediction Decile

In Table 16, entries are the percentage of actual defaults in the industry group in the listed decile of the predicted probability distribution of bankruptcy across out-of-sample firm-months. Percentages may not sum to 100% due to rounding. Predicted probability distributions use estimates from the model listed at the top of each panel. Out-of-sample predictions are from the period 2001–2008, while predictions are generated using data from 1987–2000. "ROC" is the area under the receiver operating characteristic curve.

Decile	Dilution	Kitchen Sink	Random Effects	TL/TA, $1/\sigma_E$	CHS	BIC5
Panel A. All Firms						
1	65.00%	59.06%	64.38%	62.81%	64.69%	64.69%
2	13.13%	15.94%	12.50%	13.50%	13.13%	12.50%
3	5.94%	5.94%	6.25%	6.88%	5.94%	6.88%
4	4.38%	4.69%	4.38%	4.38%	4.38%	2.81%
5	4.06%	4.06%	3.13%	4.06%	2.81%	4.69%
6–10	7.50%	10.31%	9.38%	8.38%	9.06%	8.44%
ROC	0.865	0.839	0.860	0.848	0.858	0.862
Panel B. Manufacturing						
1	72.00%	58.00%	68.00%	52.67%	64.00%	68.67%
2	12.67%	16.00%	16.67%	16.00%	20.00%	15.33%
3	5.33%	5.33%	7.33%	12.67%	8.00%	8.00%
4	3.33%	8.00%	1.33%	6.67%	3.33%	0.67%
5	0.67%	2.67%	0.00%	4.67%	0.67%	1.33%
6–10	6.00%	10.00%	6.67%	7.34%	4.00%	6.00%
ROC	0.895	0.847	0.888	0.838	0.858	0.889
Panel C. Transportation						
1	46.00%	34.00%	42.00%	42.00%	50.00%	52.00%
2	22.00%	30.00%	18.00%	18.00%	20.00%	20.00%
3	16.00%	14.00%	14.00%	14.00%	8.00%	12.00%
4	10.00%	10.00%	8.00%	8.00%	6.00%	2.00%
5	2.00%	6.00%	6.00%	6.00%	0.00%	2.00%
6–10	4.00%	6.00%	12.00%	12.00%	16.00%	12.00%
ROC	0.820	0.791	0.802	0.800	0.777	0.815
Panel D. Retail						
1	52.00%	28.00%	48.00%	28.00%	48.00%	52.00%
2	16.00%	16.00%	12.00%	16.00%	12.00%	20.00%
3	16.00%	12.00%	12.00%	20.00%	12.00%	4.00%
4	4.00%	4.00%	8.00%	8.00%	8.00%	16.00%
5	0.00%	12.00%	4.00%	8.00%	4.00%	0.00%
6-10	12.00%	28.00%	16.00%	20.00%	16.00%	8.00%
ROC	0.837	0.703	0.790	0.702	0.787	0.822
Panel E. Service						
1	65.00%	53.33%	45.00%	60.67%	60.67%	48.33%
2	10.00%	6.67%	8.33%	10.00%	16.67%	15.00%
3	6.67%	6.67%	8.33%	3.33%	1.67%	1.67%
4	1.67%	6.67%	6.67%	7.00%	1.67%	3.33%
5	3.33%	8.33%	3.33%	3.67%	5.00%	1.67%
6-10	13.33%	18.33%	28.33%	15.33%	14.33%	30.00%
ROC	0.828	0.763	0.696	0.785	0.801	0.723

variables to a hazard model including only accounting variables.[25] The ROC scores also show that model averaging improves performance. Gains in the ROC score relative to the kitchen sink model range from 0.024 to 0.119, with larger gains related to less misclassification in the bottom half of the bankruptcy decile

[25]Comparing results using the dilution prior and kitchen sink models in Panel A of Tables 15 and 16 to Table 7 in Shumway ((2001), p. 122), all report an increase of approximately 6 percentage points in the number of firms classified in the top decile when using the new approach. The improvement in the bottom half of the distribution is also comparable, with nearly one-quarter of the firms classified in the bottom half moving into the upper half of the predicted range under the new approach.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

distribution. Model-averaged forecasts also consistently outperform the CHS and BIC5 models, with gains in the ROC score between 0.005 and 0.099.

Overall, the decile analysis shows that the per-firm gains of 1.2% to 5.1% in default forecast accuracy as measured by predictive log scoring are comparable to the large gains in forecast accuracy found in previous studies. These gains are associated with improvements in ROC and decile predictions. Thus, Bayesian model averaging can generate significant increases in out-of-sample forecast accuracy.

VI. Conclusion

I develop a Bayesian model-averaging approach for predicting firm bankruptcies and defaults at a 12-month horizon using hazard models. I compare the model-averaged out-of-sample forecasts to those of popular single models and compare the variables determined to be significant predictors of default by both procedures. I find that model averaging has out-of-sample performance superior to other common models and that accounting for model uncertainty identifies far fewer variables as strong predictors of either event than conventional approaches. These results are robust to the use of different priors over the model space, including a dilution prior that weights estimates based on the correlations between the variables included in each model.

Of the 19 variables under consideration, only the ratio of total liabilities to total assets and the inverse volatility of market equity are robustly correlated with default in the overall sample and every industry group, showing that these two variables have the most empirical support for inclusion in firm default models. I also find that the data recommend excluding nine popular default predictors from such models, with seven of those excluded in the majority of individual industry groups as well. In out-of-sample forecasting, model averaging yields more accurate forecasts of future bankruptcies and defaults than several benchmark models that can be made from the 19 covariates. I find that the forecast gains from model averaging are of significant magnitude and comparable to the forecast gains made by adding market variables to bankruptcy models. I also find that including industry effects to allow for exchangeability uncertainty can improve out-of-sample bankruptcy and default forecasts, in contrast to previous results in the literature.

This paper shows that failing to consider model uncertainty leads to understating the uncertainty surrounding estimated default probabilities, which has implications for research into how investors respond to default risk. The methods used here for applying Bayesian model averaging also work well in other settings. The Laplace approximation technique allows model averaging to be applied at much lower computational cost in such areas as corporate mergers, project investment decisions, or hiring choices when the decision is discrete. Dilution priors are a flexible way to incorporate information about the composition of the correct model of default when potential covariates are highly correlated, as is the case in many economic problems. The Bayesian model-averaging techniques detailed here can also be used to inform the functional form assumptions used in modeling in a wide variety of areas, including splines and alternative likelihood functions, as discussed in the context of bankruptcy prediction by Giordani et al. (2014).

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

The finding that only the ratio of total liabilities to total assets and the inverse volatility of market equity are robust correlates of bankruptcy and default suggests that empirical work to refine individual predictors should focus on these variables to generate improvements in predictive power that are also robust to model uncertainty concerns. Overall, Bayesian model averaging offers researchers a way to incorporate empirically important concerns about model uncertainty and make gains in out-of-sample forecasting accuracy.

Appendix. Model Details

1. Computing Maximum Likelihood Estimation Approximation to Bayesian Model Averaging

I estimate model parameters using a standard multiperiod logit over the pooled firm-month observations. The parameter estimates and variance–covariance matrix of a multiperiod logit estimated in this way are identical to those of a discrete-period hazard model.

Formally, I estimate

$$P_{t-l}(y_{it} = 1) = \frac{1}{1+e^{-\beta x_{i,t-l}}}, \tag{A-1}$$

where y_{it} is an indicator equal to 1 if firm i declares bankruptcy in t, and $x_{i,t-l}$ is a set of explanatory variables observable in month $t-l$. The vector of parameters β is then estimated by maximizing the log-likelihood function

$$L(\beta) = \sum_i \sum_t y_{it} \times \ln\left(\frac{1}{1+e^{-\beta x_{i,t-l}}}\right) + (1-y_{it}) \times \ln\left(\frac{e^{-\beta x_{i,t-l}}}{1+e^{-\beta x_{i,t-l}}}\right).$$

I use the inverse of the Fisher information matrix as the base estimate of the variance–covariance matrix. I also use the Huber–White variance estimator to correct for heteroskedasticity, and I cluster the variance within firms. This technique requires only the assumption that observations across firms are independent, whereas observations within a firm over time are allowed to have an arbitrary correlation structure.

The clustered variance estimator $\hat{V}_C$ is given by

$$\hat{V}_C = \hat{V}\left(\frac{K}{K-1}\sum_{k=1}^{K} u^{k'}u^k\right)\hat{V}, \tag{A-2}$$

where K is the total number of firms in the data, k is a specific firm,

$$\hat{V} = \left(\frac{\partial^2 L(\beta)}{\partial \beta^2}\right)^{-1}$$

is the inverse of the Fisher information matrix, and

$$u^k = \sum_{j \in k} u_j,$$

where

$$u_j = \left(y_j \times \frac{f(\beta x_j)}{F(\beta x_j)} - (1-y_j) \times \frac{f(\beta x_j)}{1-F(\beta x_j)}\right) \times x_j$$

is the score of observation j, the first derivative of the log likelihood with respect to β. $F(\cdot)$ is the logit cumulative distribution function (CDF) given in equation (A-1), and $f(\cdot)$ is the logit probability distribution function (PDF). I use equations (A-1) and (A-2) to estimate parameter values and variances for all possible combinations of explanatory variables.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

2. Computing Model Fit Measures

$P(y|m)$ represents the goodness of fit of model m relative to all the other models in the model space. Raftery (1995) and Volinsky et al. (1996) show that the BIC can be used to approximate $P(y|m)$ when $\hat{\beta}_m$ is estimated via maximum likelihood, as in a discrete time hazard model. As such,

$$\text{BIC}_m \quad = \quad -2 \times L_m + k \times \ln(n),$$

where L_m is the log likelihood of model m, k is the number of parameters estimated in model m, and n is the total number of observations. The BIC is then used to approximate $P(y|m)$ as

$$P(y|m) \quad \approx \quad \exp\left(-\frac{1}{2} \times \text{BIC}_m\right),$$

and applying this approximation yields

$$P(m|y) \quad \approx \quad \frac{P(m) \times \exp\left(-\frac{1}{2} \times \text{BIC}_m\right)}{\sum_{m \in M} P(m) \times \exp\left(-\frac{1}{2} \times \text{BIC}_m\right)} \tag{A-3}$$

as the posterior probability of model m given data y, the desired weights to use in summing the individual model parameter estimates $\hat{\beta}_m$ in equation (1).

Some researchers have criticized the use of the BIC as a measure of model fit in equation (A-3) by pointing out that the $\ln(n)$ term captures only the total number of observations in the sample and not their distribution across cases. Weakliem (1999) and the comments following summarize this debate. This criticism may be applicable in this paper: Default is a rare event for a firm, and so the number of censored firms far outweighs the number of bankrupt or defaulted firms. To address this, I also calculate results using the AIC as an approximation for $P(y|m)$. AIC is given by

$$\text{AIC}_m \quad = \quad -2 \times L_m + 2k,$$

and it thus lacks the $\ln(n)$ term. I then compute an alternative version of $P(y|m)$ by using AIC_m in place of BIC_m in equation (A-3).

3. Laplace Regularity of Posterior Likelihood Function

Recall that to use the Laplace approximation for the posterior likelihood function, the likelihood function must be Laplace regular. Crawford (1994) shows that any finite mixture of exponential family distributions is Laplace regular as long as the parameters of the distributions are identifiable. Thus, it suffices to show that the integrand in equation (3) is an exponential family distribution.

Because the maximum of a function does not change under a positive monotonic transformation, I consider the exponentiated form of equation (4), which is the integrand in equation (3). I then rewrite the distributions for the prior and data separately to show that each of them is a member of the exponential family of distributions. A distribution is a member of the exponential family if it can be written in the form

$$f(z|\theta,\phi) \quad = \quad \exp\left[\frac{z\theta - b(\theta)}{a(\phi)} + c(z,\phi)\right],$$

where $a(\phi)$, $b(\theta)$, and $c(z,\phi)$ are functions of their respective parameters. I can then obtain the prior likelihood $f(\beta_m|m)$ up to a constant by setting $z = y_{it}$, $\theta = \beta_m x_{mit}$, $\phi = g(x'_{mit} \times x_{mit})^{-1}$, $a(\phi) = \phi$, $b(\theta) = \theta^2/2$, and $c(z,\phi) = -1/2\ln(2\pi\phi) - (z^2/2\phi)$. Similarly,

I obtain the data likelihood $f(y|\beta_m, m)$ by setting $z = y_{it}$, $\theta = \beta_m x_{mit}$, $\phi = 1$, $a(\phi) = \phi$, $b(\theta) = \ln\left(1 + e^{-\theta}\right)$, and $c(z, \phi) = 0$. Because both the prior and data distributions are members of the exponential family, their product is also a member of the exponential family. Thus, the likelihood function in equation (4) is Laplace regular, so the Laplace approximation to equation (3) has the properties stated in the text.

References

Altman, E. I. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance*, 23 (1968), 589–609.

Avramov, D. "Stock Return Predictability and Model Uncertainty." *Journal of Financial Economics*, 64 (2002), 423–458.

Azevedo-Filho, A., and R. D. Shachter. "Laplace's Method Approximations for Probabilistic Inference in Belief Networks with Continuous Variables." In *Uncertainty in Artificial Intelligence*, R. L. de Mantaras and D. Poole, eds. San Francisco, CA: Morgan Kaufmann (1994), 24–36.

Bharath, S. T., and T. Shumway. "Forecasting Default with the Merton Distance to Default Model." *Review of Financial Studies*, 21 (2008), 1339–1369.

Brock, W., and S. N. Durlauf. "Growth Empirics and Reality." *World Bank Economic Review*, 15 (2001), 229–272.

Brock, W. A.; S. N. Durlauf; and K. D. West. "Policy Evaluation in Uncertain Economic Environments." *Brookings Papers on Economic Activity*, 1 (2003), 235–322.

Campbell, J. Y.; J. Hilscher; and J. Szilagyi. "In Search of Distress Risk." *Journal of Finance*, 63 (2008), 2899–2939.

Chava, S., and R. A. Jarrow. "Bankruptcy Prediction with Industry Effects." *Review of Finance*, 8 (2004), 537–569.

Cohen, R. B.; C. Polk; and T. Vuolteenaho. "The Value Spread." *Journal of Finance*, 58 (2003), 609–641.

Crain, B. R., and R. L. Morgan. "Asymptotic Normality of the Posterior Distribution for Exponential Models." *Annals of Statistics*, 3 (1975), 223–227.

Crawford, S. L. "An Application of the Laplace Method to Finite Mixture Distributions." *Journal of the American Statistical Association*, 89 (1994), 259–267.

Cremers, K. J. M. "Stock Return Predictability: A Bayesian Model Selection Perspective." *Review of Financial Studies*, 15 (2002), 1223–1249.

Crosbie, P., and J. Bohn. "Modeling Default Risk." Working Paper, Moody's KMV (2003).

Dichev, I. D. "Is the Risk of Bankruptcy a Systematic Risk?" *Journal of Finance*, 53 (1998), 1131–1147.

Duffie, D.; A. Eckner; G. Horel; and L. Saita. "Frailty Correlated Default." *Journal of Finance*, 64 (2009), 2089–2123.

Duffie, D.; L. Saita; and K. Wang. "Multi-Period Corporate Default Prediction with Stochastic Covariates." *Journal of Financial Economics*, 83 (2007), 635–665.

Durlauf, S. N.; P. Johnson; and J. Temple. "Growth Econometrics." In *Handbook of Economic Growth*, P. Aghion and S. Durlauf, eds. Amsterdam, Netherlands: North-Holland (2005).

Durlauf, S. N.; A. Kourtellos; and C. M. Tan. "Are Any Growth Theories Robust?" *Economic Journal*, 118 (2008), 329–346.

Fernandez, C.; E. Ley; and M. F. J. Steel. "Benchmark Priors for Bayesian Model Averaging." *Journal of Econometrics*, 100 (2001a), 381–427.

Fernandez, C.; E. Ley; and M. F. J. Steel. "Model Uncertainty in Cross-Country Growth Regressions." *Journal of Applied Econometrics*, 16 (2001b), 563–576.

Freedman, D. A. "Diagnostics Cannot Have Much Power Against General Alternatives." *International Journal of Forecasting*, 25 (2009), 833–839.

George, E. I. "Discussion of 'Bayesian Model Averaging and Model Search' by Merlise Clyde." *Bayesian Statistics*, 6 (1999), 175–177.

Giordani, P.; T. Jacobson; E. von Schedvin; and M. Villani. "Taking the Twists into Account: Predicting Firm Bankruptcy Risk with Splines of Financial Ratios." *Journal of Financial and Quantitative Analysis*, 49 (2014), 1071–1099.

Harada, K.; T. Ito; and S. Takahashi. "Is the Distance to Default a Good Measure in Predicting Bank Failures? Case Studies." NBER Working Paper No. 16182 (2010).

Hillegeist, S. A.; E. K. Keating; D. P. Cram; and K. G. Lundstedt. "Assessing the Probability of Bankruptcy." *Review of Accounting Studies*, 9 (2004), 5–34.

Jeffreys, H. *Theory of Probability*. Oxford, UK: Oxford University Press (1961).

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press

Kandel, S., and R. F. Stambaugh. "On the Predictability of Stock Returns: An Asset Allocation Perspective." *Journal of Finance*, 51 (1996), 385–424.
Kass, R. E., and A. E. Raftery. "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion." *Journal of the American Statistical Association*, 90 (1995), 928–934.
Kass, R. E.; L. Tierney; and J. B. Kadane. "The Validity of Posterior Expansions Based on Laplace's Method." In *Bayesian and Likelihood Methods in Statistics and Econometrics*, S. Geissner, J. S. Hodges, S. J. Press, and A. Zellner, eds. Amsterdam, Netherlands: North-Holland (1990), 473–488.
Koop, G., and S. Potter. "Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging." Federal Reserve Bank of New York Staff Report 163 (2003).
Leamer, E. E. *Specification Searches*. New York, NY: John Wiley & Sons (1978).
Ley, E., and M. F. J. Steel. "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression." *Journal of Applied Econometrics*, 24 (2009), 651–674.
Merton, R. C. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance*, 29 (1974), 449–470.
Ohlson, J. A. "Financial Ratios and the Probabilistic Prediction of Bankruptcy." *Journal of Accounting Research*, 18 (1980), 109–131.
Pastor, L. "Portfolio Selection and Asset Pricing Models." *Journal of Finance*, 55 (2000), 179–223.
Pastor, L., and R. F. Stambaugh. "Costs of Equity Capital and Model Mispricing." *Journal of Finance*, 54 (1999), 67–121.
Pastor, L., and R. F. Stambaugh. "Comparing Asset Pricing Models: An Investment Perspective." *Journal of Financial Economics*, 56 (2000), 335–381.
Raftery, A. E. "Bayesian Model Selection in Social Research." *Sociological Methodology*, 25 (1995), 111–163.
Raftery, A. E. "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models." *Biometrika*, 83 (1996), 251–266.
Rosenkranz, S. L.; A. E. Raftery; and P. Diehr. "Covariate Selection in Hierarchical Models of Hospital Admission Counts: A Bayes Factor Approach." Technical Report No. 268, University of Washington (1994).
Sala-i-Martin, X.; G. Doppelhofer; and R. I. Miller. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review*, 94 (2004), 813–835.
Schervish, M. J. "A General Method for Comparing Probability Assessors." *Annals of Statistics*, 17 (1989), 1856–1879.
Shumway, T. "Forecasting Bankruptcy More Accurately: A Simple Hazard Model." *Journal of Business*, 74 (2001), 101–124.
Stock, J. H., and M. W. Watson. "An Empirical Comparison of Methods for Forecasting Using Many Predictors." Working Paper, Harvard University (2005).
Tierney, L., and J. B. Kadane. "Approximations for Posterior Moments and Marginal Densities." *Journal of the American Statistical Association*, 81 (1986), 82–86.
Tierney, L.; R. E. Kass; and J. B. Kadane. "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions." *Journal of the American Statistical Association*, 84 (1989), 710–716.
Topaloglu, Z., and Y. Yildirim. "Bankruptcy Prediction." Working Paper, Georgetown University (2009).
Vassalou, M., and Y. Xing. "Default Risk in Equity Returns." *Journal of Finance*, 59 (2004), 831–868.
Volinsky, C. T.; D. Madigan; A. E. Raftery; and R. A. Kronmal. "Bayesian Model Averaging in Proportional Hazard Models: Assessing Stroke Risk." Technical Report No. 302, University of Washington (1996).
Weakliem, D. L. "A Critique of the Bayesian Information Criterion for Model Selection." *Sociological Methods and Research*, 27 (1999), 359–397.
Winkler, R. L. "Scoring Rules and the Evaluation of Probability Assessors." *Journal of the American Statistical Association*, 64 (1969), 1073–1078.
Wright, J. H. "Bayesian Model Averaging and Exchange Rate Forecasts." *Journal of Econometrics*, 146 (2008), 329–341.
Wright, J. H. "Forecasting U.S. Inflation by Bayesian Model Averaging." *Journal of Forecasting*, 28 (2009), 131–144.
Zellner, A. "On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions." In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, P. K. Goel and A. Zellner, eds. Amsterdam, Netherlands: Elsevier Science (1986), 233–243.
Zimjewski, M. E. "Methodological Issues Related to the Estimation of Financial Distress Prediction Models." *Journal of Accounting Research*, 22 (1984), 59–82.

https://doi.org/10.1017/S002210901700031X Published online by Cambridge University Press